

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2003, 09:32:00 ; Search time 99 Seconds
(without alignments)
3164.366 Million cell updates/sec

Title: NP_000123

Sequence: 1 mpeletstcfcilicfcsa.....wvhgalmeylvcgaagly 2351

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_101002:*

- 1: /SID2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
- 2: /SID2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
- 3: /SID2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
- 4: /SID2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
- 5: /SID2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
- 6: /SID2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
- 7: /SID2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
- 8: /SID2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
- 9: /SID2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
- 10: /SID2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
- 11: /SID2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
- 12: /SID2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
- 13: /SID2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
- 14: /SID2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
- 15: /SID2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
- 16: /SID2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
- 17: /SID2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
- 18: /SID2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
- 19: /SID2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
- 20: /SID2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
- 21: /SID2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
- 22: /SID2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
- 23: /SID2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Query Length	DB ID	Description
1	12418	100.0	2351	6 AAP50059	Human factor VIII.
2	12418	100.0	2351	18 AAM11329	Human factor VIII
3	12418	100.0	2351	19 AAM46245	Human factor VII P
4	12418	100.0	2351	19 AAM44373	Human factor VIII.
5	12418	100.0	2351	21 AAY52537	Human full-length
6	12414	100.0	2351	18 AAM11435	Active Factor VIII
7	12414	100.0	2351	18 AAM11416	Active Factor VIII
8	12414	100.0	2351	18 AAM11343	Active Factor VIII
9	12413	100.0	2351	18 AAM11461	Active Factor VIII
10	12413	100.0	2351	18 AAM11445	Active Factor VIII

11	12413	100.0	2351	18 AAM11425	Active Factor VIII
12	12413	100.0	2351	18 AAM11419	Active Factor VIII
13	12413	100.0	2351	18 AAM11398	Active Factor VIII
14	12413	100.0	2351	18 AAM11387	Active Factor VIII
15	12413	100.0	2351	18 AAM11362	Active Factor VIII
16	12412	100.0	2351	9 AAP81113	Factor VIII encode
17	12412	100.0	2351	9 AAP80659	Sequence of human
18	12412	100.0	2351	18 AAM11371	Active Factor VIII
19	12412	100.0	2351	18 AAM11427	Active Factor VIII
20	12412	100.0	2351	18 AAM11408	Active Factor VIII
21	12412	100.0	2351	18 AAM11347	Active Factor VIII
22	12412	100.0	2351	18 AAM11332	Active Factor VIII
23	12411	99.9	2351	18 AAM11396	Active Factor VIII
24	12411	99.9	2351	18 AAM11371	Active Factor VIII
25	12409	99.9	2351	17 AAM00465	Factor-VIII. Homo
26	12409	99.9	2351	18 AAM11454	Active Factor VIII
27	12409	99.9	2351	22 AAB48643	Human factor VIII,
28	12408	99.9	2351	18 AAM11399	Active Factor VIII
29	12408	99.9	2351	18 AAM11404	Active Factor VIII
30	12407.5	99.9	2352	18 AAM11456	Active Factor VIII
31	12407.5	99.9	2352	18 AAM11458	Active Factor VIII
32	12407.5	99.9	2352	18 AAM11459	Active Factor VIII
33	12407.5	99.9	2352	18 AAM11463	Active Factor VIII
34	12407.5	99.9	2352	18 AAM11464	Active Factor VIII
35	12407.5	99.9	2352	18 AAM11438	Active Factor VIII
36	12407.5	99.9	2352	18 AAM11439	Active Factor VIII
37	12407.5	99.9	2352	18 AAM11442	Active Factor VIII
38	12407.5	99.9	2352	18 AAM11450	Active Factor VIII
39	12407.5	99.9	2352	18 AAM11451	Active Factor VIII
40	12407.5	99.9	2352	18 AAM11453	Active Factor VIII
41	12407.5	99.9	2352	18 AAM11454	Active Factor VIII
42	12407.5	99.9	2352	18 AAM11456	Active Factor VIII
43	12407.5	99.9	2352	18 AAM11429	Active Factor VIII
44	12407.5	99.9	2352	18 AAM11433	Active Factor VIII
45	12407.5	99.9	2352	18 AAM11403	Active Factor VIII

ALIGNMENTS

RESULT 1
ID AAP50059 standard; protein: 2351 AA.
AC AAP50059;
XX
XX 27-OCT-1991 (first entry)
XX
XX Human factor VIII.
DE Human factor VIII.
KW Human factor VIII;
XX
OS Homo sapiens.
XX
XX EP160457-A.
XX
XX 06-NOV-1985.
XX
XX 18-APR-1985; 85EP-0302734.
XX
XX 20-APR-1984; 84US-0602312.
XX
XX (GETH) GENENTECH INC.
XX
XX Capon DJ, Vohar GA, Lawn RM, Wood WJ;
XX WPI: 1985-27976/45.
XX N-PSDB: AAN50054.
XX
XX New recombinant functional human factor VIII or deriv. - useful for
XX treating haemophilia and obtd. as pure prod. by recombinant DNA
XX technology.

PS Disclosure; Fig. 10a(10-10C(III); 109bp; English.

XX The sequence is that of human factor VIII. Amino acids 1-19 are the
CC predicted signal peptide, and amino acids 1-2332 are the predicted
CC mature protein. The protein is produced in pure form and in useful
CC amounts, using recombinant DNA techniques. Factor VIII can be used
CC to correct factor VIII deficient plasma, and activates factor X to
CC Xa in the presence of factor IXa, Ca and phospholipid. These
CC activities are inactivated by antibodies specific for factor VIII.
CC The activity of the prod is bound to an immobilised monoclonal
CC antibody specific for factor VIII. Factor VIII activity is
CC activated by thrombin. The activity is bound to, and can be eluted
CC from, immobilised von Willebrand factor. Dose of factor VIII is
CC 20-40 units/kg over 8 hours i.v. for maintenance therapy for
CC haemophilia. 40 units/kg for preoperative conditions, or 15-20
CC units/kg for minor haemorrhaging.

SQ Sequence 2351 AA;

Query Match 100.0%; Score 12418; Db 6; Length 2351;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 2351; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
1 MOETSTCFPCILRPFCSATRRYTGAVELSMYQSDGLGELPDARPPRRVKSPPFN 60
1 MOETSTCFPCILRPFCSATRRYTGAVELSMYQSDGLGELPDARPPRRVKSPPFN 60
61 TSYYKKTLFEFTDHLFNIAKRPWMGLGPTIOAEYDVYITIKNASHVSIHAN 120
61 TSYYKKTLFEFTDHLFNIAKRPWMGLGPTIOAEYDVYITIKNASHVSIHAN 120
61 TSYYKKTLFEFTDHLFNIAKRPWMGLGPTIOAEYDVYITIKNASHVSIHAN 120
121 GVSYKASBGAAYDQTSQREKEDVPFGSHYVWYAKENGPASDPLCTLYSLH 180
121 GVSYKASBGAAYDQTSQREKEDVPFGSHYVWYAKENGPASDPLCTLYSLH 180
121 GVSYKASBGAAYDQTSQREKEDVPFGSHYVWYAKENGPASDPLCTLYSLH 180
121 GVSYKASBGAAYDQTSQREKEDVPFGSHYVWYAKENGPASDPLCTLYSLH 180
181 VDLKDLNSGLIGALVLCRGSLAKKQVTLAKTLLFANFDEGKSMSETRKSLMDRD 240
181 VDLKDLNSGLIGALVLCRGSLAKKQVTLAKTLLFANFDEGKSMSETRKSLMDRD 240
181 VDLKDLNSGLIGALVLCRGSLAKKQVTLAKTLLFANFDEGKSMSETRKSLMDRD 240
241 AASARAMPKMTVNGVYNSLPGLIGCHKRSYVHWITGMCTTPEVHSIFLEGHTFLVRH 300
241 AASARAMPKMTVNGVYNSLPGLIGCHKRSYVHWITGMCTTPEVHSIFLEGHTFLVRH 300
241 AASARAMPKMTVNGVYNSLPGLIGCHKRSYVHWITGMCTTPEVHSIFLEGHTFLVRH 300
301 RQASLEISPIETFLAOTLMDLGOFLIFCHLISHQHDGMEAYKVYVSCPEPOLRMKNNE 360
301 RQASLEISPIETFLAOTLMDLGOFLIFCHLISHQHDGMEAYKVYVSCPEPOLRMKNNE 360
301 RQASLEISPIETFLAOTLMDLGOFLIFCHLISHQHDGMEAYKVYVSCPEPOLRMKNNE 360
361 EADYDDDLTJSEMDVYRFPDDNSPFIQISYAKKHPKTWVHYIAEBEEMDAPLVLA 420
361 EADYDDDLTJSEMDVYRFPDDNSPFIQISYAKKHPKTWVHYIAEBEEMDAPLVLA 420
361 EADYDDDLTJSEMDVYRFPDDNSPFIQISYAKKHPKTWVHYIAEBEEMDAPLVLA 420
421 PDDRSTKSOYLANNGPQRIKRYKAYFMAVYDEFTKREAIQHESGILGPILYGVEDTL 480
421 PDDRSTKSOYLANNGPQRIKRYKAYFMAVYDEFTKREAIQHESGILGPILYGVEDTL 480
421 PDDRSTKSOYLANNGPQRIKRYKAYFMAVYDEFTKREAIQHESGILGPILYGVEDTL 480
481 LILFNQASRPYNIYPGITDVRPLXSRRLRGVYKHLKDPILLGSEIRKYKVTVEDGP 540
481 LILFNQASRPYNIYPGITDVRPLXSRRLRGVYKHLKDPILLGSEIRKYKVTVEDGP 540
481 LILFNQASRPYNIYPGITDVRPLXSRRLRGVYKHLKDPILLGSEIRKYKVTVEDGP 540
541 TKSDPCLTRYYSSVNMERDLASGLIGPLILCYKESVDQGNQIMSDKRNVIYFSVDE 600
541 TKSDPCLTRYYSSVNMERDLASGLIGPLILCYKESVDQGNQIMSDKRNVIYFSVDE 600
541 TKSDPCLTRYYSSVNMERDLASGLIGPLILCYKESVDQGNQIMSDKRNVIYFSVDE 600
601 NRSWYLTENIORELPNPAQVLEDPFOASNTMISINGYFDSLQSYCLHEVYVWYILS 660
601 NRSWYLTENIORELPNPAQVLEDPFOASNTMISINGYFDSLQSYCLHEVYVWYILS 660
601 NRSWYLTENIORELPNPAQVLEDPFOASNTMISINGYFDSLQSYCLHEVYVWYILS 660
661 IGAQDELSPFSGYTRKHVYEDLTLFPGSGYVPMENMGALICGNSDFRNKG 720
661 IGAQDELSPFSGYTRKHVYEDLTLFPGSGYVPMENMGALICGNSDFRNKG 720
661 IGAQDELSPFSGYTRKHVYEDLTLFPGSGYVPMENMGALICGNSDFRNKG 720
721 MTAALVSSCDKNTGDIYEDSYEDISAVYLKNNALIEPNSFSONSHRSTROKQFNATTI 780
721 MTAALVSSCDKNTGDIYEDSYEDSYEDISAVYLKNNALIEPNSFSONSHRSTROKQFNATTI 780
721 MTAALVSSCDKNTGDIYEDSYEDSYEDISAVYLKNNALIEPNSFSONSHRSTROKQFNATTI 780
```

```
721 MTAALVSSCDKNTGDIYEDSYEDSYEDISAVYLKNNALIEPNSFSONSHRSTROKQFNATTI 780
721 MTAALVSSCDKNTGDIYEDSYEDSYEDISAVYLKNNALIEPNSFSONSHRSTROKQFNATTI 780
781 PENDIKETDPMFAHRTMPKRIQVSSDILLMLRQSPPTPGLSLSDLOEAYEFESDDPS 840
781 PENDIKETDPMFAHRTMPKRIQVSSDILLMLRQSPPTPGLSLSDLOEAYEFESDDPS 840
781 PENDIKETDPMFAHRTMPKRIQVSSDILLMLRQSPPTPGLSLSDLOEAYEFESDDPS 840
841 PGADIDNNISLPMTHFRPOLAHSGDVPFPESGLOLANKGTTAAATELKDFKVSST 900
841 PGADIDNNISLPMTHFRPOLAHSGDVPFPESGLOLANKGTTAAATELKDFKVSST 900
841 PGADIDNNISLPMTHFRPOLAHSGDVPFPESGLOLANKGTTAAATELKDFKVSST 900
901 SNNLITITSDMLAAGDNTSSIGPSPMYHSDOLTLPFGKSPPLTESGGRSLSE 960
901 SNNLITITSDMLAAGDNTSSIGPSPMYHSDOLTLPFGKSPPLTESGGRSLSE 960
901 SNNLITITSDMLAAGDNTSSIGPSPMYHSDOLTLPFGKSPPLTESGGRSLSE 960
961 NNDKILLESGLMNSOESSMGRNVSTESGRLEFKGRRAHGALLTDNALFVYSISLKTIN 1020
961 NNDKILLESGLMNSOESSMGRNVSTESGRLEFKGRRAHGALLTDNALFVYSISLKTIN 1020
961 NNDKILLESGLMNSOESSMGRNVSTESGRLEFKGRRAHGALLTDNALFVYSISLKTIN 1020
1021 KTSNNSATNRKTHIDGPSLLIENSPPVNONLIESDTEFKKVTPLIHDRMLDKNAVALRL 1080
1021 KTSNNSATNRKTHIDGPSLLIENSPPVNONLIESDTEFKKVTPLIHDRMLDKNAVALRL 1080
1021 KTSNNSATNRKTHIDGPSLLIENSPPVNONLIESDTEFKKVTPLIHDRMLDKNAVALRL 1080
1081 NMSNKTTSKMMEMVQOKKEGIPPDANPDMSEFFMLPLPESARWIORTHGKNSLNSG 1140
1081 NMSNKTTSKMMEMVQOKKEGIPPDANPDMSEFFMLPLPESARWIORTHGKNSLNSG 1140
1081 NMSNKTTSKMMEMVQOKKEGIPPDANPDMSEFFMLPLPESARWIORTHGKNSLNSG 1140
1141 QGSPKQVSLGPEKSYEGONFLSEKNKYVKGDEFTKDVGLKEMVFPSSRNPLTLMDN 1200
1141 QGSPKQVSLGPEKSYEGONFLSEKNKYVKGDEFTKDVGLKEMVFPSSRNPLTLMDN 1200
1141 QGSPKQVSLGPEKSYEGONFLSEKNKYVKGDEFTKDVGLKEMVFPSSRNPLTLMDN 1200
1201 LHMNNHNEKKEIOEIEKEETLQENYVLPQIRHYVGTAKNPKNPLFLSTRONVESYD 1260
1201 LHMNNHNEKKEIOEIEKEETLQENYVLPQIRHYVGTAKNPKNPLFLSTRONVESYD 1260
1201 LHMNNHNEKKEIOEIEKEETLQENYVLPQIRHYVGTAKNPKNPLFLSTRONVESYD 1260
1261 GATAPVLDQFNSLNDSTNRKTAFAFSKGEEMZGLGNQKQVEXYACTRTSPN 1320
1261 GATAPVLDQFNSLNDSTNRKTAFAFSKGEEMZGLGNQKQVEXYACTRTSPN 1320
1261 GATAPVLDQFNSLNDSTNRKTAFAFSKGEEMZGLGNQKQVEXYACTRTSPN 1320
1321 SOONFYORSKALKQFRLPLEETLEKKIYDDTSTOWSKNNKHLTPSTLQIDYENKE 1380
1321 SOONFYORSKALKQFRLPLEETLEKKIYDDTSTOWSKNNKHLTPSTLQIDYENKE 1380
1321 SOONFYORSKALKQFRLPLEETLEKKIYDDTSTOWSKNNKHLTPSTLQIDYENKE 1380
1381 KGAITOSPDLCTRSHSIPQANRSLPLAKVSFSPSIRPYLTRVLEODNSSHLPAASY 1440
1381 KGAITOSPDLCTRSHSIPQANRSLPLAKVSFSPSIRPYLTRVLEODNSSHLPAASY 1440
1381 KGAITOSPDLCTRSHSIPQANRSLPLAKVSFSPSIRPYLTRVLEODNSSHLPAASY 1440
1441 RKDQSGVQESSHFTLOGAKNNLSIALITLGMTGQDREVSIGTSATNSVYKKEVNTLP 1500
1441 RKDQSGVQESSHFTLOGAKNNLSIALITLGMTGQDREVSIGTSATNSVYKKEVNTLP 1500
1441 RKDQSGVQESSHFTLOGAKNNLSIALITLGMTGQDREVSIGTSATNSVYKKEVNTLP 1500
1501 KPDLPTSGKVELLPKYHIYOKDLFPTESNGSGCHLDVSGSLQGTGCAIKYENARP 1560
1501 KPDLPTSGKVELLPKYHIYOKDLFPTESNGSGCHLDVSGSLQGTGCAIKYENARP 1560
1501 KPDLPTSGKVELLPKYHIYOKDLFPTESNGSGCHLDVSGSLQGTGCAIKYENARP 1560
1561 GKVPPLVATESSAKTPSKLLDPLANDNHYGTQIPKEBMSQSEKTAFFKKDTILST 1620
1561 GKVPPLVATESSAKTPSKLLDPLANDNHYGTQIPKEBMSQSEKTAFFKKDTILST 1620
1561 GKVPPLVATESSAKTPSKLLDPLANDNHYGTQIPKEBMSQSEKTAFFKKDTILST 1620
1621 NACESNHAIAINEGONKPELEYMAKQRTRELCSONPVLKRHRDREITRTTLOSQOE 1680
1621 NACESNHAIAINEGONKPELEYMAKQRTRELCSONPVLKRHRDREITRTTLOSQOE 1680
1621 NACESNHAIAINEGONKPELEYMAKQRTRELCSONPVLKRHRDREITRTTLOSQOE 1680
1681 IDYDDITSEYMKKEDFDIYDEDEQSPFSQFKTTHFPIAAYRLMDYMSSSPYVLRNR 1740
1681 IDYDDITSEYMKKEDFDIYDEDEQSPFSQFKTTHFPIAAYRLMDYMSSSPYVLRNR 1740
1681 IDYDDITSEYMKKEDFDIYDEDEQSPFSQFKTTHFPIAAYRLMDYMSSSPYVLRNR 1740
1741 AAGSGVPOKRVYPOEFDQSPQPLYRBELENEHIGLGYTAEEVENDIMVTPPNQSR 1800
1741 AAGSGVPOKRVYPOEFDQSPQPLYRBELENEHIGLGYTAEEVENDIMVTPPNQSR 1800
1741 AAGSGVPOKRVYPOEFDQSPQPLYRBELENEHIGLGYTAEEVENDIMVTPPNQSR 1800
1801 PYSFYSLSLSTIEDOROGAEPRKNTKPEMETTYTWKVOHHMAPTEDEDCANAYTSDV 1860
1801 PYSFYSLSLSTIEDOROGAEPRKNTKPEMETTYTWKVOHHMAPTEDEDCANAYTSDV 1860
1801 PYSFYSLSLSTIEDOROGAEPRKNTKPEMETTYTWKVOHHMAPTEDEDCANAYTSDV 1860
```

```

QY 1861 DLEKDVHSGILGRLVCHNTLTNPAHGRQVYQEFALFTIPDETQSWTFENNERCRA 1920
XX |||
XX 1861 DLEKDVHSGILGRLVCHNTLTNPAHGRQVYQEFALFTIPDETQSWTFENNERCRA 1920
Db |||
QY 1921 PCNIQMEDPTREKYNRRHAINGIYMDTLPGIYVMAODORIRWYLLSMGSENIHSIHSGH 1980
XX |||
XX 1921 PCNIQMEDPTREKYNRRHAINGIYMDTLPGIYVMAODORIRWYLLSMGSENIHSIHSGH 1980
Db |||
QY 1981 VFTVKKKEEYKMAIYNLYPGVFETVYMLPESKAGIARVECLIGHLHAGNSTLFLYVSNKC 2040
XX |||
XX 1981 VFTVKKKEEYKMAIYNLYPGVFETVYMLPESKAGIARVECLIGHLHAGNSTLFLYVSNKC 2040
Db |||
QY 2041 QTEPLGMAHGIRDFQITASGOYQOMAPKLARLHYSGSINAMSTKEPFSWIKVDLLAPMI 2100
XX |||
XX 2041 QTEPLGMAHGIRDFQITASGOYQOMAPKLARLHYSGSINAMSTKEPFSWIKVDLLAPMI 2100
Db |||
QY 2101 HGIKTQAGAROKFSSLYISOFILMYSLDGGKMQYRGNSGTGLAVFGANDSSGSIKHNIEN 2160
XX |||
XX 2101 HGIKTQAGAROKFSSLYISOFILMYSLDGGKMQYRGNSGTGLAVFGANDSSGSIKHNIEN 2160
Db |||
QY 2161 PRITARYIRLHPTHTYRSRTLRLMELMGCDLNSCSWPLGMSKAIISDAQTASSTYTNMFA 2220
XX |||
XX 2161 PRITARYIRLHPTHTYRSRTLRLMELMGCDLNSCSWPLGMSKAIISDAQTASSTYTNMFA 2220
Db |||
QY 2221 TNSPSKARLHLQGSNMRPOVNNPKREKQVNDQOKMKTGYGTQGYKSLTSMYKREPL 2280
XX |||
XX 2221 TNSPSKARLHLQGSNMRPOVNNPKREKQVNDQOKMKTGYGTQGYKSLTSMYKREPL 2280
Db |||
QY 2281 ISSSQDGHQWTLFQNGKRVYFQGNDSFTPVVNSLDPILLIRLRIHPQSVHOLALRM 2340
XX |||
XX 2281 ISSSQDGHQWTLFQNGKRVYFQGNDSFTPVVNSLDPILLIRLRIHPQSVHOLALRM 2340
Db |||
QY 2341 EVLGCENADLY 2351
XX |||
XX 2341 EVLGCENADLY 2351
Db |||

```

```

PR 11-JUL-1995: 95US-0001025.
XX
XX (CHIR ) CHIRON CORP.
XX
XX Cohen FE, Hung DT, Innis M;
XX
XX WPI: 1997-119050/11.
XX
XX N-PSDB: A0513357.
XX
XX Factor VIII:C analog modified adjacent to a non-activating Arg
XX residue - used in the treatment of haemophilias, by improvement of
XX haemostasis
XX
XX Disclosure: Fig 1: 90pp: English.
XX
XX This sequence represents the native Factor VIII:C. Factor VIII:C is a
XX large glycoprotein that participates in the blood coagulation cascade
XX that ultimately converts soluble fibrinogen to insoluble fibrin clot,
XX effecting haemostasis. A deficiency in Factor VIII:C is responsible for
XX haemophilia A, which is an X-chromosome-linked inherited bleeding
XX diathesis. Factor VIII:C is activated by plasma proteases, such as
XX thrombin. During activation the mature polypeptide is cleaved to
XX generate heavy and light chain fragments that are further cleaved. The
XX DNA encoding this sequence is mutated, using mutagenic primers, to
XX produce the active Factor VIII:C analogues of the invention (such as
XX AAH11329). The analogues comprise a native Factor VIII:C polypeptide
XX modified at a site adjacent to a non-activating Arg residue so that a
XX Arg-Pro or Pro-Arg dipeptide is created. Complexes of two or more of the
XX analogues, nucleic acids and vectors encoding them may be used alone or
XX in conjunction with each other, for the prevention or treatment of active
XX Factor VIII:C deficiency in a mammal. The analogues may be used as
XX immunogens to raise antibodies, and in the treatment of haemophilias, by
XX improvement of haemostasis. The analogues are resistant to proteolytic
XX cleavage and display increased plasma half-life. They may be administered
XX at lower dosages and by different modes of administration.
XX
XX Sequence 2351 AA:
XX
XX Query Match 100.0%; Score 12418; DB 18; Length 2351;
XX Best Local Similarity 100.00%; Pred. No. 0;
XX Matches 2351; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 481 LIIFFNQASREYNTYPRGHTDVRAPLYSRRLPKGVKHLKDFLLPEIRFKMTVTEDG 540
 DB 481 LIIFFNQASREYNTYPRGHTDVRAPLYSRRLPKGVKHLKDFLLPEIRFKMTVTEDG 540
 QY 541 TSDPRCLTRYSSSFVMMERDLASGLIGPLIYCYESVDQNGQIMSKRRVILFVSFDE 600
 DB 541 TSDPRCLTRYSSSFVMMERDLASGLIGPLIYCYESVDQNGQIMSKRRVILFVSFDE 600
 QY 601 NRSWTLFENIORPLPNAGVQLEDEPEFQASNMHMSINGVYFSDLOASCLHEVAVYIIS 660
 DB 601 NRSWTLFENIORPLPNAGVQLEDEPEFQASNMHMSINGVYFSDLOASCLHEVAVYIIS 660
 QY 661 IGADDFLSVFSGYFQKHKMYEDTLTFPGEYFPMSEMPGATIGCHNSDPFRNG 720
 DB 661 IGADDFLSVFSGYFQKHKMYEDTLTFPGEYFPMSEMPGATIGCHNSDPFRNG 720
 QY 721 MTALLKVVSCDKNTGYEDSYEDISAVLUSKNNAIIEPFSQNSRHPSTQKOPNATTI 780
 DB 721 MTALLKVVSCDKNTGYEDSYEDISAVLUSKNNAIIEPFSQNSRHPSTQKOPNATTI 780
 QY 781 PENDIEKTDPMFAHRTPMKQIONVSSDLMILNQSPPHGLSLSDLOAKYETFSDDPS 840
 DB 781 PENDIEKTDPMFAHRTPMKQIONVSSDLMILNQSPPHGLSLSDLOAKYETFSDDPS 840
 QY 841 PGADSNNSLSEMTHERPOLHSGDVFTPESGQLRLNKLGTATATLKKLDFKVSST 900
 DB 841 PGADSNNSLSEMTHERPOLHSGDVFTPESGQLRLNKLGTATATLKKLDFKVSST 900
 QY 901 SNMLISTIPDNLAAGDNTSLGPPSPVHYDSQDPTTFGKSSPLTFESGGLSSEE 960
 DB 901 SNMLISTIPDNLAAGDNTSLGPPSPVHYDSQDPTTFGKSSPLTFESGGLSSEE 960
 QY 961 NDSKILLESGLMNSQESSWGNVSTESGRLEKGRAGHALLTKONALEKYSISLKTN 1020
 DB 961 NDSKILLESGLMNSQESSWGNVSTESGRLEKGRAGHALLTKONALEKYSISLKTN 1020
 QY 1021 KTSNNSATNRKTHIDGSLIENSFSWONLIESDTEFKVTPLIHDRMLDKNATLRL 1080
 DB 1021 KTSNNSATNRKTHIDGSLIENSFSWONLIESDTEFKVTPLIHDRMLDKNATLRL 1080
 QY 1081 NMSNKTSSSKNMEMVOQKEGPIPPDAONPDMSEFKMLFLPSASAWIORTHGKNSLNG 1140
 DB 1081 NMSNKTSSSKNMEMVOQKEGPIPPDAONPDMSEFKMLFLPSASAWIORTHGKNSLNG 1140
 QY 1141 QGSPKQVSLGPEKSYEGONFLSEKNKYVGKGEFTKDGLEKEMFSPSRMLFLTNLDN 1200
 DB 1141 QGSPKQVSLGPEKSYEGONFLSEKNKYVGKGEFTKDGLEKEMFSPSRMLFLTNLDN 1200
 QY 1201 LHEHNTHNOEKKIOBEIEKKEELLIOENVVLPQIHVTGTGKFNKMLFLSTQONVSGSYD 1260
 DB 1201 LHEHNTHNOEKKIOBEIEKKEELLIOENVVLPQIHVTGTGKFNKMLFLSTQONVSGSYD 1260
 QY 1261 GAYAPVLODFRSLNDSTNRKHTAHFSKSGEENLEGLNOTKQIYERYACTTRISPT 1320
 DB 1261 GAYAPVLODFRSLNDSTNRKHTAHFSKSGEENLEGLNOTKQIYERYACTTRISPT 1320
 QY 1321 SOONFVTOXSKRALKOPRLEETELERLIYDDOSTYOMSKNNKHLPTSTLQIDYNEKE 1380
 DB 1321 SOONFVTOXSKRALKOPRLEETELERLIYDDOSTYOMSKNNKHLPTSTLQIDYNEKE 1380
 QY 1381 KGAITOSPLODCLTRSHSIPQANRSPPLAKVSSPSPTRPIYLRVLPQDNSSHPAASY 1440
 DB 1381 KGAITOSPLODCLTRSHSIPQANRSPPLAKVSSPSPTRPIYLRVLPQDNSSHPAASY 1440
 QY 1441 RKDQSGVDESSHFLQAGAKNNLSLALILEMTGDORREVSLGTSATNSVYKRVENVLP 1500
 DB 1441 RKDQSGVDESSHFLQAGAKNNLSLALILEMTGDORREVSLGTSATNSVYKRVENVLP 1500
 QY 1501 KPDLPTSGKVELLPKHYIYQNDLPFTLESNSPCHDLIVGSLGCTGCAIKMEANRP 1560
 DB 1501 KPDLPTSGKVELLPKHYIYQNDLPFTLESNSPCHDLIVGSLGCTGCAIKMEANRP 1560

QY 1561 GKVPFLVATESSAKTBSKLIADPLANDNHGQIPIREKMSQOEKSPKTAFFKKTIIISL 1620
 DB 1561 GKVPFLVATESSAKTBSKLIADPLANDNHGQIPIREKMSQOEKSPKTAFFKKTIIISL 1620
 QY 1621 NACBSNHAIAAINEQONKPEIEVYMAKQOTERLCSQNPPLKRRHOREITRTTLOSDEE 1680
 DB 1621 NACBSNHAIAAINEQONKPEIEVYMAKQOTERLCSQNPPLKRRHOREITRTTLOSDEE 1680
 QY 1681 IDYDDTISVEMAKKEDFDIYDEDEQSPRSFQKTRHRYIAVERLMDYGMSSPVLNRR 1740
 DB 1681 IDYDDTISVEMAKKEDFDIYDEDEQSPRSFQKTRHRYIAVERLMDYGMSSPVLNRR 1740
 QY 1741 AOGSGVPOFKKVVPEFTDGSFTQPIRGLENEHGLGPIYIRAEVDNNTMTFNRQSR 1800
 DB 1741 AOGSGVPOFKKVVPEFTDGSFTQPIRGLENEHGLGPIYIRAEVDNNTMTFNRQSR 1800
 QY 1801 PYSFSSLSIYEDDQROGABERKMYKPMETKYFMYQOHMAPTNDEPCKRAMAFSDV 1860
 DB 1801 PYSFSSLSIYEDDQROGABERKMYKPMETKYFMYQOHMAPTNDEPCKRAMAFSDV 1860
 QY 1861 DLEKDVHSLIGPLVCHTNTLPNABGROYVQEFALFTTIDETKSWTFENNERCRA 1920
 DB 1861 DLEKDVHSLIGPLVCHTNTLPNABGROYVQEFALFTTIDETKSWTFENNERCRA 1920
 QY 1921 PCNIOMEDPTFKENYRFAHNGYIMOTLPGLVMAODORIRMYLLSMGSNENIHSIHFSQH 1980
 DB 1921 PCNIOMEDPTFKENYRFAHNGYIMOTLPGLVMAODORIRMYLLSMGSNENIHSIHFSQH 1980
 QY 1981 VFTYRKEEYKNAIYNYPGVEFVEMLPKAGIMRVPCILGEBLHAGMSTFLIYSNKC 2040
 DB 1981 VFTYRKEEYKNAIYNYPGVEFVEMLPKAGIMRVPCILGEBLHAGMSTFLIYSNKC 2040
 QY 2041 QPIGMAHGHIRDFOTASGOYQMAPKLARLHYSISINASTKEPSSATKVDLAPMTI 2100
 DB 2041 QPIGMAHGHIRDFOTASGOYQMAPKLARLHYSISINASTKEPSSATKVDLAPMTI 2100
 QY 2101 HGKTOGAROKFESSIYISOFTIMTSLDGKKMOTYRGNSTGLTAVFPGVDSGIRKINIFN 2160
 DB 2101 HGKTOGAROKFESSIYISOFTIMTSLDGKKMOTYRGNSTGLTAVFPGVDSGIRKINIFN 2160
 QY 2161 PPIIARIYRLHPHYISIRSLRMEIMCDSLNSCMLPGMESALISDAQIYASSYPTNMA 2220
 DB 2161 PPIIARIYRLHPHYISIRSLRMEIMCDSLNSCMLPGMESALISDAQIYASSYPTNMA 2220
 QY 2221 TMSPSKARLHLOGRSNAMPPOVNNPKMLQVDOQTKMYKTGVTQGVKSLTSMYKKEFL 2280
 DB 2221 TMSPSKARLHLOGRSNAMPPOVNNPKMLQVDOQTKMYKTGVTQGVKSLTSMYKKEFL 2280
 QY 2281 ISSQODHOMTLTFQNGKRYKPOGNOSFPPVNSIDPPLIRYIRLHPOKWHOTAIAPM 2340
 DB 2281 ISSQODHOMTLTFQNGKRYKPOGNOSFPPVNSIDPPLIRYIRLHPOKWHOTAIAPM 2340
 QY 2341 EYLCEAODLY 2351
 DB 2341 EYLCEAODLY 2351

RESULT 3
 AA#6245 standard; Protein; 2351 AA.
 ID AA#6245;
 AC AA#6245;
 XX
 DT 06-AUG-1998 (first entry)
 XX
 DE Human factor III protein sequence.
 XX
 KW Replication defective; recombinant retrovirus; RRV; therapeutic protein; hemophilia; thrombosis; hypercoagulable disorder; liver disease; human; hepatitis; thalassemia; phenylketonuria; Leisch-Nyhan syndrome; diabetes; cystic fibrosis; Duchenne's muscular dystrophy; hypercholesterolemia; hypopituitarism; adenine deaminase deficiency; HIV infection; anemia; Guacher's syndrome; high blood pressure; Alzheimer's disease; factor III;

KM	autoimmune; inflammatory disease
OS	Homo sapiens.
PN	MO9800541-N2.
XX	08-JAN-1998.
PD	02-JUL-1997; 97NO-DS11784.
PF	04-JUN-1997; 97US-0869309.
PR	03-JUN-1996; 96US-0645601.
PR	13-NOV-1996; 96US-0696381.
XX	(CHIR) CHIRON CORP.
PA	Allen JR, Barber JR, Boder M, Chang SM, Chong K;
PI	De LA VEGA D, Depoloni J, Greengard J, Hsu DC, Ibanez CE;
PI	Jolly D, Lee R, Mittelstaedt DM, Pirvasak CE, Respass JG;
XX	WPI; 1998-086966/08.
DR	N-PSDB; AAV19580.
XX	New replication defective recombinant retro-viruses - which can be
PT	administered to provide long term systemic expression of therapeutic
PT	protein in blood, useful in, e.g. treating hyper-coagulable
PT	disorders
XX	Example 28; Pages 203-210; 272pp; English.
XX	This is the human factor III sequence. The encoding DNA is used in the
CC	construction of recombinant retroviral vectors expressing human
CC	factor VIII. The specification provides the preparation of replication
CC	defective recombinant retrovirus (RVV) expressing a therapeutic protein.
CC	The RVV preparation is resistant to degradation by human complement and
CC	is capable of inducing long term systemic expression of the therapeutic
CC	protein when administered intravenously to a human. The long term
CC	systemic depression results in a measurable level of the therapeutic
CC	protein being produced in the blood of the human for a period of at least
CC	30 days after the administration of the RVV vector preparation. RVV's can
CC	be used for in vivo delivery of therapeutic protein to treat, e.g.,
CC	hemophilia A, hemophilia B, thrombosis, hypercoagulable disorders,
CC	liver diseases such as hepatitis, disorders such as thalassemia,
CC	phenylketonuria, Lesch-Nyhan syndrome, severe combined immunodeficiency
CC	(SCID), cystic fibrosis, Duchenne's Muscular Dystrophy, inherited
CC	emphysema, familial hypercholesterolemia, diabetes, hypopituitarism,
CC	adenine deaminase deficiency, alpha-1-antitrypsin deficiency, Gaucher's
CC	syndrome, anaemia, infections such as HIV infection, high blood pressure,
CC	Alzheimer's disease, autoimmune or inflammatory disease or graft versus
CC	host disease. RVV's are capable of surviving inactivation in human serum
CC	thereby allowing efficient gene transfer over prolonged periods of time.
XX	
SQ	Sequence 2351 AA;
Query Match	100.0%; Score 12418; DB 19; Length 2351;
Best Local Similarity	100.0%; Pred. No. 0;
Matches 2351;	Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	1 MOEILSTCFPLCLRCFCSATRRYLVANFLSMYDMSQSLGFLPDAFPPRYKSPFN 60
DB	1 MOEILSTCFPLCLRCFCSATRRYLVANFLSMYDMSQSLGFLPDAFPPRYKSPFN 60
QY	61 TSVYVKKTLVETEDHLEFINAFPRPMWGLGPTIOAEYVDTVITLKNASHPVSLAV 120
DB	61 TSVYVKKTLVETEDHLEFINAFPRPMWGLGPTIOAEYVDTVITLKNASHPVSLAV 120
QY	121 GVSYTWKASBGALEYDDQTSQREKEDKVPFGGSHTYVWYLKENGPMASDPLCTIYSYLH 180
DB	121 GVSYTWKASBGALEYDDQTSQREKEDKVPFGGSHTYVWYLKENGPMASDPLCTIYSYLH 180
QY	181 VDLVKDINSGLIGALLVCRGSLAKETQTLHAFILFLFVPEBGKSWHSETKSLNQDSD 240
DB	181 VDLVKDINSGLIGALLVCRGSLAKETQTLHAFILFLFVPEBGKSWHSETKSLNQDSD 240

Qy	241	AASRARPKNHYNQVNRSLPGLIGHRSKYVWYHVGCTTPVHSHIFLECHTFYLVNH	300
Dh	241	AASRARPKNHYNQVNRSLPGLIGHRSKYVWYHVGCTTPVHSHIFLECHTFYLVNH	300
Qy	301	ROASLEISPTTFLTAQTLMLMDGQPLFCITSHQHDGMEAYVYDSCPEEPOLRKKNNE	360
Dh	301	ROASLEISPTTFLTAQTLMLMDGQPLFCITSHQHDGMEAYVYDSCPEEPOLRKKNNE	360
Qy	361	EAEYDDDLTDSEMDVYREDDNSPSTQIRSAVKKHPRKTVWYHIAAEEDMDYAPLVLA	420
Dh	361	EAEYDDDLTDSEMDVYREDDNSPSTQIRSAVKKHPRKTVWYHIAAEEDMDYAPLVLA	420
Qy	421	PDDSSYSQYLNNGPORIGRKYKRVFMAATDTEFTKRALQIHESGILGELYGEGDTL	480
Dh	421	PDDSSYSQYLNNGPORIGRKYKRVFMAATDTEFTKRALQIHESGILGELYGEGDTL	480
Qy	481	LIFRNQASAPRYTYPHGJTDVPRYPSNRPLPKVVKLKNOPLLDGEIFFXKKMYVEDOP	540
Dh	481	LIFRNQASAPRYTYPHGJTDVPRYPSNRPLPKVVKLKNOPLLDGEIFFXKKMYVEDOP	540
Qy	541	TKSDPRCLTYRYSFPMNERDLAGLIGPLLTCKYSVDQGNQIMSDKNVLYLSEYDE	600
Dh	541	TKSDPRCLTYRYSFPMNERDLAGLIGPLLTCKYSVDQGNQIMSDKNVLYLSEYDE	600
Qy	601	NRSWYLTENTQRELPNPACVOLDEPFOASNIMHSINGYEDSLQLSVCLAEVAYWYILS	660
Dh	601	NRSWYLTENTQRELPNPACVOLDEPFOASNIMHSINGYEDSLQLSVCLAEVAYWYILS	660
Qy	661	IGAQDDELFSYFSGGYFFKHMYEDTLTLPPFSGEYFMSMEHNGMLTLCCHSDPRNG	720
Dh	661	IGAQDDELFSYFSGGYFFKHMYEDTLTLPPFSGEYFMSMEHNGMLTLCCHSDPRNG	720
Qy	721	MYALIKYSCDKMTGTYGYDSEDIQSAVYLTKSKNNALEPSTSONSRHPSTROKOFNATI	780
Dh	721	MYALIKYSCDKMTGTYGYDSEDIQSAVYLTKSKNNALEPSTSONSRHPSTROKOFNATI	780
Qy	781	PENDERTKDPMFAHPRKPMKIONVSSDMLLRQSPYTHGJLSJDQLEAKYETFSDDPS	840
Dh	781	PENDERTKDPMFAHPRKPMKIONVSSDMLLRQSPYTHGJLSJDQLEAKYETFSDDPS	840
Qy	841	PGALDSNNLSLEKTHPRPOLJHSGDMVFPESGLQULRNEKLGTYAATBLKLDKFRVYST	900
Dh	841	PGALDSNNLSLEKTHPRPOLJHSGDMVFPESGLQULRNEKLGTYAATBLKLDKFRVYST	900
Qy	901	NNNLISTIPSDNLAAQTDWTSISLAPPSPMAYHYSQDUTLLEFKSSPILTESGGPLSLE	960
Dh	901	NNNLISTIPSDNLAAQTDWTSISLAPPSPMAYHYSQDUTLLEFKSSPILTESGGPLSLE	960
Qy	961	NNDSKILLESGLMSSOGSSGKNVSTESGRLPFGKRAHBPALLTFDNALRKVYSILKTN	1020
Dh	961	NNDSKILLESGLMSSOGSSGKNVSTESGRLPFGKRAHBPALLTFDNALRKVYSILKTN	1020
Qy	1021	KTSNKSATNRKTHIDGPSLLBNSPSWQNILSDPERFKVTPPLIHDMIMDKNATATRL	1080
Dh	1021	KTSNKSATNRKTHIDGPSLLBNSPSWQNILSDPERFKVTPPLIHDMIMDKNATATRL	1080
Qy	1081	NHMSNKTTSKNNEMVQOKKEGPIPPDAONPDMSFFKMLFLEPSARMTOITHOKNSLNG	1140
Dh	1081	NHMSNKTTSKNNEMVQOKKEGPIPPDAONPDMSFFKMLFLEPSARMTOITHOKNSLNG	1140
Qy	1141	OGPSRKOIYVGLGPEKVEBOONFLSKNNVYVVGGETTKOYGLKEAVPSSRNPLFLNND	1200
Dh	1141	OGPSRKOIYVGLGPEKVEBOONFLSKNNVYVVGGETTKOYGLKEAVPSSRNPLFLNND	1200
Qy	1201	LHNENTHNOCKTQOEIEKKETLQOENWVLPOLHYTGYCKNNKMLLELSTRONVGSST	1260
Dh	1201	LHNENTHNOCKTQOEIEKKETLQOENWVLPOLHYTGYCKNNKMLLELSTRONVGSST	1260
Qy	1261	GAYAVVLQDPRSLNDSTNRKTHAHFSKKGEBENLEGLANOTKOIYEKTACTTIRISPT	1320
Dh	1261	GAYAVVLQDPRSLNDSTNRKTHAHFSKKGEBENLEGLANOTKOIYEKTACTTIRISPT	1320

```

QY 1321 SQQNFVQSRKALQKQRLPLEETELEKRIIVDTOSTQSKMKHLPSTLTQIDYNEKE 1380
DB 1322 SQQNFVQSRKALQKQRLPLEETELEKRIIVDTOSTQSKMKHLPSTLTQIDYNEKE 1380
QY 1381 KGAITQSPSLDCLTRSHSIPQANRSPPLIAKVSFSPISPIYLTFLVLPQDNSSHLPAASY 1440
DB 1381 KGAITQSPSLDCLTRSHSIPQANRSPPLIAKVSFSPISPIYLTFLVLPQDNSSHLPAASY 1440
QY 1441 RKKDSGVQESSHFLQCAKKNKNSLAILTLEMTGQDQFVSGISATNSVTKYKYEVLVP 1500
DB 1441 RKKDSGVQESSHFLQCAKKNKNSLAILTLEMTGQDQFVSGISATNSVTKYKYEVLVP 1500
QY 1501 KPDLPRKTSKVELLPKHVHYOKDLFPTETSGSGHLDLVESGLQGTGKAIKNEANRP 1560
DB 1501 KPDLPRKTSKVELLPKHVHYOKDLFPTETSGSGHLDLVESGLQGTGKAIKNEANRP 1560
QY 1561 GVPPLRVATSSAKTPSKLDPAMDNNHGTQIPKEEMKSOEKSPKTAARKKDTLLST 1620
DB 1561 GVPPLRVATSSAKTPSKLDPAMDNNHGTQIPKEEMKSOEKSPKTAARKKDTLLST 1620
QY 1621 NACESNHAIAINEGONKPEIEVWAKQRTERLCSQNPVILKRHORERTTLQSDQEE 1680
DB 1621 NACESNHAIAINEGONKPEIEVWAKQRTERLCSQNPVILKRHORERTTLQSDQEE 1680
QY 1681 IDYDPTISEMKKEDPDIYDEENQSPRSFQKTRHYFIAVBRMDYGMSSPHVLNR 1740
DB 1681 IDYDPTISEMKKEDPDIYDEENQSPRSFQKTRHYFIAVBRMDYGMSSPHVLNR 1740
QY 1741 AOSGVPQFKVYQOETGOSTFQPLRYGELNEHLGLGYTIAVEENIMVTERQNASR 1800
DB 1741 AOSGVPQFKVYQOETGOSTFQPLRYGELNEHLGLGYTIAVEENIMVTERQNASR 1800
QY 1801 PYSFYSLSISTEEDQOGAPRKPNKYNKTKYTFKVOHMAPTKDEFDCKANAYSDV 1860
DB 1801 PYSFYSLSISTEEDQOGAPRKPNKYNKTKYTFKVOHMAPTKDEFDCKANAYSDV 1860
QY 1861 DLEKDVHSGILGPLVCHNTLNPAGROVYQOEFALFFIEDKTSWYFTENNERCRA 1920
DB 1861 DLEKDVHSGILGPLVCHNTLNPAGROVYQOEFALFFIEDKTSWYFTENNERCRA 1920
QY 1921 PCNIOMEDPTFKENRYRHAINGYINDTLPGVMAQDORIMWYLLSGMSNENHSHFSGH 1980
DB 1921 PCNIOMEDPTFKENRYRHAINGYINDTLPGVMAQDORIMWYLLSGMSNENHSHFSGH 1980
QY 1981 VFTVARKKEEYKMAIYNTPGVEFTYEMKPSRAGITWRKJIGSHIAGKSTLFTYNSKC 2040
DB 1981 VFTVARKKEEYKMAIYNTPGVEFTYEMKPSRAGITWRKJIGSHIAGKSTLFTYNSKC 2040
QY 2041 QTPPLGASGHIRDQITASQOYQOMAPKLARLHSGSINAMSTKEPFSWIKVDLAPMI 2100
DB 2041 QTPPLGASGHIRDQITASQOYQOMAPKLARLHSGSINAMSTKEPFSWIKVDLAPMI 2100
QY 2101 HGIKTOGARQKRFSSLYTSOFTIMYSLDGKKQWYRGNSGTGLVWFGVNDSSGKIHNFN 2160
DB 2101 HGIKTOGARQKRFSSLYTSOFTIMYSLDGKKQWYRGNSGTGLVWFGVNDSSGKIHNFN 2160
QY 2161 PPIIARYIRLHPTHYRSIRSLRMLKGLCDLNSCSPGLGMSKASISDAQITASSYFTNFA 2220
DB 2161 PPIIARYIRLHPTHYRSIRSLRMLKGLCDLNSCSPGLGMSKASISDAQITASSYFTNFA 2220
QY 2221 TWSPSKARLHLOGRSMARPOVNNPKEMLOVPOFKTMKVTGVTQGVKSLNSMYVEFL 2280
DB 2221 TWSPSKARLHLOGRSMARPOVNNPKEMLOVPOFKTMKVTGVTQGVKSLNSMYVEFL 2280
QY 2281 ISSSDQGHOMTLFQNGKAVYFQGNDSFTPVVNSLDBPLLTRLRIBHQSVMHQAIALRM 2340
DB 2281 ISSSDQGHOMTLFQNGKAVYFQGNDSFTPVVNSLDBPLLTRLRIBHQSVMHQAIALRM 2340
QY 2341 EVLGCEADOLY 2351
DB 2341 EVLGCEADOLY 2351

```

```

RESULT 4
AAM44373
ID AAM44373 standard; Protein: 2351 AA.
XX
AC AAM44373;
XX
DT 20-JUL-1998 (first entry)
XX
DE Human Factor VIII.
XX
KW Factor VIII; blood clotting; haemophilia A; gene therapy;
KW retrovirus; vector; human.
XX
OS Homo sapiens.
XX
FH Key
FT Domain
FT /note="B-domain"
XX
PN WO9800542-A2.
XX
PD 08-JAN-1998.
XX
PE 02-JUL-1997; 97MO-US11785.
XX
PR 04-JUN-1997; 97US-0869309.
PR 03-JUL-1996; 96US-0645601.
PR 13-AUG-1996; 96US-0696381.
XX
PA (CHIR ) CHIRON CORP.
XX
PI Allen JR, Barber JR, Boder M, Chang SM, Chong K;
PI De la Vega D, Depolo NJ, Greengard J, Hsu DC, Ibanez CE;
PI Jolly DJ, Mittlestaedt DM, Prusak CE, Reespeck JG;
DR MPI: 1998-086967/08.
DR N-PSDB; AAV15359.
XX
PT New replication defective recombinant retroviruses - which express B
PT domain-deleted human factor VIII or human factor IX for the
PT treatment of haemophiliaa
XX
PS Example 28; Page 164-166; 236pp; English.
XX
CC This polypeptide comprises human Factor VIII. The invention
CC relates to preparations of replication-defective recombinant
CC retrovirus (RV) expressing a B domain-deleted human Factor VIII
CC protein, such as the SQN deletion mutant (see AAM44372), where the
CC recombinant RV is capable of infecting human cells, is resistant
CC to degradation by human complement and is capable of inducing
CC long-term (at least 30 days and up to 6 months or longer
CC post-injection) systemic expression of Factor VIII when
CC administered to a haemophilia A patient.
XX
SQ Sequence 2351 AA:
Query Match 100.0%; Score 12418; DB 19; Length 2351;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2351; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MOIELSTCFPLCLLRCEGATRRRYLAVALSDMYQOSDGLPVDARPPRPKSFPPN 60
DB 1 MOIELSTCFPLCLLRCEGATRRRYLAVALSDMYQOSDGLPVDARPPRPKSFPPN 60
QY 61 TSVYRKTLVEVTHDLRNLAKPRPPMGLGPTIOAEYIDIVYITLKNASHPVSLHAY 120
DB 61 TSVYRKTLVEVTHDLRNLAKPRPPMGLGPTIOAEYIDIVYITLKNASHPVSLHAY 120
QY 121 GVSYWKASGEGAEYDDOTSQREKEDKVPFGSGHTYVWOYLKENGPMASDPLCTYSYLSH 180
DB 121 GVSYWKASGEGAEYDDOTSQREKEDKVPFGSGHTYVWOYLKENGPMASDPLCTYSYLSH 180
QY 181 VDLVKNDSGLIGALLVCEGSLAKERTQTLHFKITLLFAVPEDEKSNHSEFKNSLMQDHD 240

```

181 VDLVKDNLNSGLIGALLVCREGLAKEKTOQLHAFVDEBEGSWMSSEKSLNQDDO 240
241 AASARAMPKMTYVNGYVNRSLPGLIGCHRSKYWHVIGMTPEVHSLFECHTFLVRNH 300
241 AASARAMPKMTYVNGYVNRSLPGLIGCHRSKYWHVIGMTPEVHSLFECHTFLVRNH 300
301 ROASLEISPTFLTQOTLMDLGOFLPCHTSHOHGMEAVYKDSOPEPQIRKKNNE 360
301 ROASLEISPTFLTQOTLMDLGOFLPCHTSHOHGMEAVYKDSOPEPQIRKKNNE 360
361 EABDYDDLTDSBMYVRFDDNSPFIQIRSAVKKHPKWTWVHTAAEEDMDYAPLVLA 420
361 EABDYDDLTDSBMYVRFDDNSPFIQIRSAVKKHPKWTWVHTAAEEDMDYAPLVLA 420
421 PDDRSYKSOYLNNGFORIGRKYKRYRMATYDEFTKREAIQIHESGILGPLYGEGDTL 480
421 PDDRSYKSOYLNNGFORIGRKYKRYRMATYDEFTKREAIQIHESGILGPLYGEGDTL 480
481 LIIFKNOASFPYNYIPHGITDVRPLYSRRLPKGVKHLKDFLLPGLPELTKKMTYVEDGP 540
481 LIIFKNOASFPYNYIPHGITDVRPLYSRRLPKGVKHLKDFLLPGLPELTKKMTYVEDGP 540
541 TKSDPCLTRYYSFYNMERDLASGLIGPLLYCKESVDORGNQIMSDKRNVLIFSYDE 600
541 TKSDPCLTRYYSFYNMERDLASGLIGPLLYCKESVDORGNQIMSDKRNVLIFSYDE 600
601 NRSWYLTENIORFLPAPAGVLEDEPEFOASINHSINGVYFPLQJLSYCHEVAVWYIIS 660
601 NRSWYLTENIORFLPAPAGVLEDEPEFOASINHSINGVYFPLQJLSYCHEVAVWYIIS 660
661 IGAOTDFLSYFSGYTFKKKMYEDTLTLFPESGETVPMSENPCMLTLCGHSNDRPNNG 720
661 IGAOTDFLSYFSGYTFKKKMYEDTLTLFPESGETVPMSENPCMLTLCGHSNDRPNNG 720
721 M7ALLKVSCKDKNTGDIYEDSYEDISAYLTKSKNNAIEPRSFSONSRHSTROKOFNATYI 780
721 M7ALLKVSCKDKNTGDIYEDSYEDISAYLTKSKNNAIEPRSFSONSRHSTROKOFNATYI 780
781 PENDIEKTDPMFAHRTPMPKITIONYSSSOLMLLROSPTRHGSLSDLOEAKYEFSDPS 840
781 PENDIEKTDPMFAHRTPMPKITIONYSSSOLMLLROSPTRHGSLSDLOEAKYEFSDPS 840
841 PGALDSNNSLSEMTHERPOLHSGDMVTPPESGLQJLRLNEKLGITNAATELKKLDKVSST 900
841 PGALDSNNSLSEMTHERPOLHSGDMVTPPESGLQJLRLNEKLGITNAATELKKLDKVSST 900
901 SNNLISIPSDNLAACTDNTSSLAGPPSPVHYDSQDITLFGKSSPLTESGCPLSISEE 960
901 SNNLISIPSDNLAACTDNTSSLAGPPSPVHYDSQDITLFGKSSPLTESGCPLSISEE 960
961 NNDKLTESGLMNSOESSMGKWSSTESGRLEFGKRAHGPALLTDMNLEFYVSLIKTN 1020
961 NNDKLTESGLMNSOESSMGKWSSTESGRLEFGKRAHGPALLTDMNLEFYVSLIKTN 1020
1021 KTSNNSATNRKTHIDGPSLLIENBSVWQNLISDTEFKKVPPLIHDMLMDKNATLRT 1080
1021 KTSNNSATNRKTHIDGPSLLIENBSVWQNLISDTEFKKVPPLIHDMLMDKNATLRT 1080
1081 NMSNKTSSKNNEMVQOKKEGPJPPDQNPDSFFKMLFLPESARWIOPTHGNSNSNG 1140
1081 NMSNKTSSKNNEMVQOKKEGPJPPDQNPDSFFKMLFLPESARWIOPTHGNSNSNG 1140
1081 NMSNKTSSKNNEMVQOKKEGPJPPDQNPDSFFKMLFLPESARWIOPTHGNSNSNG 1140
1141 OGPSPKOLVSLGPEKSEBQONFLSEKKNVYVGGEFTKDVGLKEVFPSSRNLFLTLMDN 1200
1141 OGPSPKOLVSLGPEKSEBQONFLSEKKNVYVGGEFTKDVGLKEVFPSSRNLFLTLMDN 1200
1201 LHENNTHOBRKIOEBIEKKEKTLIOENVVLPOIHVYTGKNNMKMLFLTLSTRONVEGSYD 1260
1201 LHENNTHOBRKIOEBIEKKEKTLIOENVVLPOIHVYTGKNNMKMLFLTLSTRONVEGSYD 1260
1261 GAVAPVLODFRSLNDSTNRKTHAFSKKEEENLEGLGNQOTKOIYKACTTRISPNF 1320
1261 GAVAPVLODFRSLNDSTNRKTHAFSKKEEENLEGLGNQOTKOIYKACTTRISPNF 1320

1261 GAVAPVLODFRSLNDSTNRKTHAFSKKEEENLEGLGNQOTKOIYKACTTRISPNF 1320
1321 SOONFYORSKRALKOFRPLFEETELEKRIIYDDTSTOWSKMKHLAPSTLQIDYNEKE 1380
1321 SOONFYORSKRALKOFRPLFEETELEKRIIYDDTSTOWSKMKHLAPSTLQIDYNEKE 1380
1381 KGATOSPLSDCLTSHSHIPOANSPLPIKVSFSPSIPITLTPRVYFONSSSHLPAASY 1440
1381 KGATOSPLSDCLTSHSHIPOANSPLPIKVSFSPSIPITLTPRVYFONSSSHLPAASY 1440
1441 RKDGSVOESSHFLQAKKNMLSLAILTLEMTDQREVSIGTSATNSVYKKEVNTVLP 1500
1441 RKDGSVOESSHFLQAKKNMLSLAILTLEMTDQREVSIGTSATNSVYKKEVNTVLP 1500
1501 KPDLPTSGVLELLPVRHYQKDLFPTETSNGSPGLDLYEGSLLOGTEGAKIKMEANRP 1560
1501 KPDLPTSGVLELLPVRHYQKDLFPTETSNGSPGLDLYEGSLLOGTEGAKIKMEANRP 1560
1561 GKVPFLVATESSATPSKLDPLAMDNHGTQIPKEEKSQEKSPERTAFKKKDTILSL 1620
1561 GKVPFLVATESSATPSKLDPLAMDNHGTQIPKEEKSQEKSPERTAFKKKDTILSL 1620
1621 NACBSNHAIAINEGONKPELEVTWAKOFTERLCSQNPVYLKRHOREITRTYLOSQOE 1680
1621 NACBSNHAIAINEGONKPELEVTWAKOFTERLCSQNPVYLKRHOREITRTYLOSQOE 1680
1681 IDYDDTISYEMKKEPDIYDEDENOSPSPROKTRHYFTAVERLMDYGSSPHYLARN 1740
1681 IDYDDTISYEMKKEPDIYDEDENOSPSPROKTRHYFTAVERLMDYGSSPHYLARN 1740
1741 AOSGSVPQFKKVVFOEFDGSTQPLRGELNHLGLCLPTRYRAVEDNITWTFRNOASR 1800
1741 AOSGSVPQFKKVVFOEFDGSTQPLRGELNHLGLCLPTRYRAVEDNITWTFRNOASR 1800
1801 PYSEYSSLSIEBEDOQOAEPRKFNVAENETKYTFKVOHNAPTKDEFDCKAMAYESDV 1860
1801 PYSEYSSLSIEBEDOQOAEPRKFNVAENETKYTFKVOHNAPTKDEFDCKAMAYESDV 1860
1861 DLEKDVHSLIGPLVCHNTLNPAGROYVQOEFALFTITDEKSKYTFENNERCRA 1920
1861 DLEKDVHSLIGPLVCHNTLNPAGROYVQOEFALFTITDEKSKYTFENNERCRA 1920
1921 PCNTOMEDPFFKFNTHFAINGYIMDTLPGLVNAODQIRMTYLSMGSNENIHSIHESGH 1980
1921 PCNTOMEDPFFKFNTHFAINGYIMDTLPGLVNAODQIRMTYLSMGSNENIHSIHESGH 1980
1981 VFTYRKKKEEKALYNTLYPGVETVEKMLPSKAGIMRVECLIGEHLHAAGSTLFLVYSNC 2040
1981 VFTYRKKKEEKALYNTLYPGVETVEKMLPSKAGIMRVECLIGEHLHAAGSTLFLVYSNC 2040
2041 QTPLGMAHGIIRFOITASGOYGOMAPRLARHYSGSINANSTREPFSTIVYDLAMPIT 2100
2041 QTPLGMAHGIIRFOITASGOYGOMAPRLARHYSGSINANSTREPFSTIVYDLAMPIT 2100
2101 HGITQOAROKFSSLIQSOFIIMTSLDKKQOYRGNSIGTLMVFEGVDSGKHNIFE 2160
2101 HGITQOAROKFSSLIQSOFIIMTSLDKKQOYRGNSIGTLMVFEGVDSGKHNIFE 2160
2161 PPIIARIYRLPHYSIRSTLMBELMGDLNCSMPLGMEKRAISDAOTITASVFTMFA 2220
2161 PPIIARIYRLPHYSIRSTLMBELMGDLNCSMPLGMEKRAISDAOTITASVFTMFA 2220
2221 TWSPSKARHLQGRSNAMRPOVNNRKEKLOYDQKTMKTYGTTQGVASLTSMTYVEFL 2280
2221 TWSPSKARHLQGRSNAMRPOVNNRKEKLOYDQKTMKTYGTTQGVASLTSMTYVEFL 2280
2281 ISSSQDQWTLFFONGKVKVQGNQDSFTPVVNSLDPLLTRYLRIHPOSVHQJALRM 2340
2281 ISSSQDQWTLFFONGKVKVQGNQDSFTPVVNSLDPLLTRYLRIHPOSVHQJALRM 2340
2341 EYLGCEAODLY 2351
2341 EYLGCEAODLY 2351

RESULT 5
 ID AAY52537 standard; Protein; 2351 AA.
 AC AAY52537;
 DT 28-FEB-2000 (first entry)
 XX Human full-length factor VIII.
 DE
 XX Factor VIII; haemophilia; proteolysis; heavy chain; light chain;
 KM secretion; von Willebrand factor; VWF; C2 domain; Intron 22; Inversion;
 KW non-functional; bleeding disorder; coagulation; treatment.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Peptide 1..19
 FT /note= "Signal peptide"
 FT Protein 20..1708
 FT /note= "Maximum length human factor VIII heavy chain"
 FT Protein 1709..2351
 FT /note= "Human factor VIII light chain"
 PN MO959622-A1.
 PD 25-NOV-1999.
 XX
 XX 17-MAY-1999; 99WO-US10872.
 XX
 XX 20-MAY-1998; 98US-0082000.
 PR
 XX (GETH) GENENTECH INC.
 PA
 PI Vohar GA;
 XX WPI: 2000-053195/04.
 DR N-PSDB; AA238604.
 DR
 XX
 XX New method for treating haemophilia A -
 FT
 PS Disclosure; Fig 9; 50pp; English.
 XX
 XX This sequence represents the full-length human factor VIII. In
 CC this form, the protein is 300 kD in size with the domain structure
 CC A1-A2-B-A3-C1-C2. However, prior to secretion, this 300 kD protein
 CC is proteolysed into a heavy chain (A1-A2-B, with continued
 CC proteolysis within the B domain resulting in molecules of varying
 CC length) and a light chain (A3-C1-C2) that remains non-covalently attached
 CC to the heavy chain. Upon secretion, factor VIII is rapidly cleared from
 CC the circulation unless it is bound by the plasma protein von Willebrand
 CC Factor (VWF). Factor VIII binds to VWF through the light chain, with
 CC known binding sites at the N-terminus and within the C2 domain.
 CC Haemophilia A is frequently caused by an intron 22 inversion in the
 CC factor VIII gene, which separates the gene into two sections, exons 1-22
 CC becoming inverted and localised telomeric to the original site, while
 CC exons 23-26 remain isolated at the original site. Exons 23-26 a portion
 CC of the C1 and all of the C2 domains, without which factor VIII is
 CC non-functional. The factor VIII gene product of individuals with this
 CC mutation thus comprises domains A1-A2-B-A3 plus a fragment of C1, which
 CC on proteolysis is non-functional, resulting in a bleeding disorder. The
 CC invention relates to a novel method for for treating haemophilia A in a
 CC mammal able to produce the factor VIII heavy chain. The method comprises
 CC administering to the mammal either an effective amount of factor VIII
 CC light chain, or a gene encoding it, and may be useful for treating
 CC patients such as those whose haemophilia A is caused by intron 22
 CC inversion. The recombinant factor VIII products of this invention are
 CC derived from well-characterised starting materials which therefore
 CC reduces the risk of pathogenic infection which was previously a problem
 CC when using donated plasma. Furthermore, the invention provides a more
 CC economic and potentially more effective treatment for haemophilia. There

CC is also a need of providing factor VIII activity to patients who
 CC produce or are at risk of producing antibodies against full-length
 CC factor VIII.
 CC
 SQ Sequence 2351 AA;
 Query Match 100.0%; Score 12418; DB 21; Length 2351;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2351; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MOTELSTCEFLCLRFCHSATRRYIYGAVALSDMDQSDGLFVDAARPPRPKSEFPN 60
 DB 1 MOTELSTCEFLCLRFCHSATRRYIYGAVALSDMDQSDGLFVDAARPPRPKSEFPN 60
 QY 61 TSVYKKTLEFEFTDHLFNIAKPRPPMGLGPTIOAEYDVTYITLKNAASHPVALAV 120
 DB 61 TSVYKKTLEFEFTDHLFNIAKPRPPMGLGPTIOAEYDVTYITLKNAASHPVALAV 120
 QY 121 GVSYKASGAIYDQTSQREKEDKTPFGSGHTYVQYVLAENGPMAADPLCLTYSLSH 180
 DB 121 GVSYKASGAIYDQTSQREKEDKTPFGSGHTYVQYVLAENGPMAADPLCLTYSLSH 180
 QY 181 VDLVKNLNGLIGALLVCREGSLAKERTQTLAKFTLLFAVDEGKSMHSETKSLMQDDSD 240
 DB 181 VDLVKNLNGLIGALLVCREGSLAKERTQTLAKFTLLFAVDEGKSMHSETKSLMQDDSD 240
 QY 241 AASARAPKMHNTVNGYVNRSLPGLIGHRKYVWHYIGCTTPEVHSTFLEGHPTLVNRH 300
 DB 241 AASARAPKMHNTVNGYVNRSLPGLIGHRKYVWHYIGCTTPEVHSTFLEGHPTLVNRH 300
 QY 301 ROASLEISPTFLTAOTLLMDLGOFLFCHISSHQHDEGAEVAVKVDSCPEPOLRMKNE 360
 DB 301 ROASLEISPTFLTAOTLLMDLGOFLFCHISSHQHDEGAEVAVKVDSCPEPOLRMKNE 360
 QY 361 EAEDYDDLDLDSRMVYRFPDDNSPFIQIRSAKHPPTVHTYIAAEEDMDVAPLVLA 420
 DB 361 EAEDYDDLDLDSRMVYRFPDDNSPFIQIRSAKHPPTVHTYIAAEEDMDVAPLVLA 420
 QY 421 PDDRSYKQYLLNNGPQIRGRKRYKRYMAVYDFEFTKRAIIOHESGILGLLYGEVDTL 480
 DB 421 PDDRSYKQYLLNNGPQIRGRKRYKRYMAVYDFEFTKRAIIOHESGILGLLYGEVDTL 480
 QY 481 LIIRKNQASRPYNIYHGITDVRPLYSRRLPKGVKHLKDFLLPEIRFYKMYTVVDEGP 540
 DB 481 LIIRKNQASRPYNIYHGITDVRPLYSRRLPKGVKHLKDFLLPEIRFYKMYTVVDEGP 540
 QY 541 TKSPDRLTRYISSFVNNMRDLASGLIPLLCYKKSVDQNRGNQMSKRVYLTFSEDE 600
 DB 541 TKSPDRLTRYISSFVNNMRDLASGLIPLLCYKKSVDQNRGNQMSKRVYLTFSEDE 600
 QY 601 NRSWYLTENIQRLPMPAGVQLEDEFEQASNMHMSINGYVFDLSLVCLEHVAWYVILS 660
 DB 601 NRSWYLTENIQRLPMPAGVQLEDEFEQASNMHMSINGYVFDLSLVCLEHVAWYVILS 660
 QY 661 IGAOTDPLSYFEGSYFKKRYEDTLTLFPFSEYVYKMEKMPGMLTIGCHNSDFNRNG 720
 DB 661 IGAOTDPLSYFEGSYFKKRYEDTLTLFPFSEYVYKMEKMPGMLTIGCHNSDFNRNG 720
 QY 721 MTALLKYSQCDKMTGDIYEDSYEDISAVILSKNNAIPRFSQNSRHPSTROKQFNATTI 780
 DB 721 MTALLKYSQCDKMTGDIYEDSYEDISAVILSKNNAIPRFSQNSRHPSTROKQFNATTI 780
 QY 781 PENDIEKTDWFAHRPMPKIONVSSDLMLLQSPTPGSLSDLOAKYEFSDPS 840
 DB 781 PENDIEKTDWFAHRPMPKIONVSSDLMLLQSPTPGSLSDLOAKYEFSDPS 840
 QY 841 PGADISNNLSLEWTHFRPOLDHSQDMVTPESGQLNLNKLGTATATLKLKIDKXVST 900
 DB 841 PGADISNNLSLEWTHFRPOLDHSQDMVTPESGQLNLNKLGTATATLKLKIDKXVST 900
 QY 901 SNNLISTPSPNLAAGDNTSSIGPSPVHYDSDLTTLFGKSSPLRESGSPISLEE 960
 DB 901 SNNLISTPSPNLAAGDNTSSIGPSPVHYDSDLTTLFGKSSPLRESGSPISLEE 960

QY	961	NNDKLLSEGLANNOSSQSMGKNVSTSEGRLEFRGRAGPALITKDNALFKVYSILDKTN	1020
Db	961	NNDKLLSEGLANNOSSQSMGKNVSTSEGRLEFRGRAGPALITKDNALFKVYSILDKTN	1020
QY	1021	KTSNSATNRKTHIDGSLIENSPLYWONILESPTREKATYPLIHBMIMDNANALTL	1080
Db	1021	KTSNSATNRKTHIDGSLIENSPLYWONILESPTREKATYPLIHBMIMDNANALTL	1080
QY	1081	NHMSNKTSSKMEVQOKKEGPIPPDAONPDMSFFKMLFPESANMIQTHGKNSLNG	1140
Db	1081	NHMSNKTSSKMEVQOKKEGPIPPDAONPDMSFFKMLFPESANMIQTHGKNSLNG	1140
QY	1141	QGPSKQVLSLGPERSVGEQNFLEKKNVYVKGGEFTKDVGLKEMVPPSSRNLFITNLDN	1200
Db	1141	QGPSKQVLSLGPERSVGEQNFLEKKNVYVKGGEFTKDVGLKEMVPPSSRNLFITNLDN	1200
QY	1201	LHENNTHNOEKIJOEIEIKKETLJOENVYLPQHHVYTGKKNPMKNLFLSTRNVGASD	1260
Db	1201	LHENNTHNOEKIJOEIEIKKETLJOENVYLPQHHVYTGKKNPMKNLFLSTRNVGASD	1260
QY	1261	GATAPVLOPRLANDSTRRTKTAHFSKSGEENLEGLNQOTQIYERKACTTRISPT	1320
Db	1261	GATAPVLOPRLANDSTRRTKTAHFSKSGEENLEGLNQOTQIYERKACTTRISPT	1320
QY	1321	SOQNFYORSKRALKQFLPLETELEKRIIVDDTSTQMSKMKHLPTSLTQIDYNEKE	1380
Db	1321	SOQNFYORSKRALKQFLPLETELEKRIIVDDTSTQMSKMKHLPTSLTQIDYNEKE	1380
QY	1381	KGATOSPLSDCLTSHSIPQANSRPLPAKVSPPSIRPIYIRVLFQONSSHLPAASY	1440
Db	1381	KGATOSPLSDCLTSHSIPQANSRPLPAKVSPPSIRPIYIRVLFQONSSHLPAASY	1440
QY	1441	RKDGVOSSHFLOGAKNNLSLAILTLEMTGDOREVSIGTSATNSVYKKEVTVLP	1500
Db	1441	RKDGVOSSHFLOGAKNNLSLAILTLEMTGDOREVSIGTSATNSVYKKEVTVLP	1500
QY	1501	KPDLPTSGKVELPRVHIYOKDLFPTETNSGSPGHLDYEGSLQGTGEGAKIMNEANRP	1560
Db	1501	KPDLPTSGKVELPRVHIYOKDLFPTETNSGSPGHLDYEGSLQGTGEGAKIMNEANRP	1560
QY	1561	GKVPFLRVATSSATPKLIDPLAMNHNHGNQIPIPEEKSOEKSPEKTFKKKDTILSL	1620
Db	1561	GKVPFLRVATSSATPKLIDPLAMNHNHGNQIPIPEEKSOEKSPEKTFKKKDTILSL	1620
QY	1621	NACESNHAIAMINEQNPPELEVTWAKQGTERRCSQNPVLKRHRORITRTLOSDOE	1680
Db	1621	NACESNHAIAMINEQNPPELEVTWAKQGTERRCSQNPVLKRHRORITRTLOSDOE	1680
QY	1681	IDYDDTISYEMKKEDEFDIYDEDEQSPRSFOKTRHYFIAAVERLMDYGMSSSPHYLRNR	1740
Db	1681	IDYDDTISYEMKKEDEFDIYDEDEQSPRSFOKTRHYFIAAVERLMDYGMSSSPHYLRNR	1740
QY	1741	AQSGVPOPKRYVROEFIDGSEFTQPLVRGBLNLHGILGAPYRAVEDNIMTFRMOASR	1800
Db	1741	AQSGVPOPKRYVROEFIDGSEFTQPLVRGBLNLHGILGAPYRAVEDNIMTFRMOASR	1800
QY	1801	PYSFTYSSLSIYEBDOQOAGPRKKNVFPNETKTYFNKVQHMAPTKDEPCKAMAYFSDV	1860
Db	1801	PYSFTYSSLSIYEBDOQOAGPRKKNVFPNETKTYFNKVQHMAPTKDEPCKAMAYFSDV	1860
QY	1861	DLEKDVHSLIGLPLVCHTNTLNPAGHQVTVQEFALFTIPDEKSKSYFEMMENENCA	1920
Db	1861	DLEKDVHSLIGLPLVCHTNTLNPAGHQVTVQEFALFTIPDEKSKSYFEMMENENCA	1920
QY	1921	PCNTOMEDPPEKENTRFPAINGYIMDTLPGVVAQOQRIRMTLLSMGSENIHSHISG	1980
Db	1921	PCNTOMEDPPEKENTRFPAINGYIMDTLPGVVAQOQRIRMTLLSMGSENIHSHISG	1980
QY	1981	VFTVRRKKEEYKALYNLYPGVFETVEMLPSKAGIMVRECLIGEHLMAGMSTLEVTYSNKC	2040
Db	1981	VFTVRRKKEEYKALYNLYPGVFETVEMLPSKAGIMVRECLIGEHLMAGMSTLEVTYSNKC	2040

QY	2041	QTPLGMAHSHIDPQITASQGYQMAPKLARLHYSGSINANSTKEPSWIKVDLLAPMII	2100
Db	2041	QTPLGMAHSHIDPQITASQGYQMAPKLARLHYSGSINANSTKEPSWIKVDLLAPMII	2100
QY	2101	HGKITGQAKQKSSSIYISQFTIMVSLDGKKNQWYNGNSGTILAMPQGNVDSGIRKHNIFN	2160
Db	2101	HGKITGQAKQKSSSIYISQFTIMVSLDGKKNQWYNGNSGTILAMPQGNVDSGIRKHNIFN	2160
QY	2161	PLIARIYRLPPTHYSIRSTRLMELMGCDLNSCSPMGESKASIDAOITASSYFTNMPA	2220
Db	2161	PLIARIYRLPPTHYSIRSTRLMELMGCDLNSCSPMGESKASIDAOITASSYFTNMPA	2220
QY	2221	TWSPSKARLHLQGRSNAMRPQVNNPKEMLDVDFOKTKMYTGVTTOGVKSILTSMYKEFL	2280
Db	2221	TWSPSKARLHLQGRSNAMRPQVNNPKEMLDVDFOKTKMYTGVTTOGVKSILTSMYKEFL	2280
QY	2281	ISSSQGHQWTLFPGNGKVKVYFQGNODSFPVYNSLDPPLTRYLRIHPQSWHOTALRM	2340
Db	2281	ISSSQGHQWTLFPGNGKVKVYFQGNODSFPVYNSLDPPLTRYLRIHPQSWHOTALRM	2340
QY	2341	EVLGCEADQDLY 2351	
Db	2341	EVLGCEADQDLY 2351	

RESULT 6	
AAW11435	
ID	AAW11435 standard; Protein; 2351 AA.
XX	
AC	AAW11435;
XX	
DT	20-NOV-1997 (first entry)
XX	
DE	Active Factor VIII:C analogue S1311X.
XX	
KW	Factor VIII:C; analogue; glycoprotein; blood coagulation cascade;
KW	fibrinogen; fibrin clot; haemostasis; haemophilia A; bleeding diathesis;
KW	plasma protease; thrombin; immunogen; antibody; haemophilic; therapy;
KW	proteolytic cleavage.
XX	
OS	Homo sapiens.
OS	Synthetic.
XX	
PH	Key
FT	Peptide
FT	Location/Qualifiers
FT	/note- "signal peptide"
FT	20..2351
FT	/note- "mature Factor VIII:C"
FT	20..1667
FT	/note- "heavy chain fragment"
FT	1330
FT	/label- Phe, Glu, Pro
FT	1668..2350
FT	/note- "light chain fragment"
FT	760..1667
FT	/note- "B domain"
XX	
XX	MO9703195-A1.
XX	
XX	30-JAN-1997.
XX	
XX	09-JUL-1996; 96MO-US11444.
XX	
XX	11-JUL-1995; 95US-0001025.
XX	
XX	(CHIR) CHIRON CORP.
XX	
XX	Cohen FE, Hung DT, Innis M;
XX	
XX	WPI; 1997-119050/11.
XX	
XX	Factor VIII:C analog modified adjacent to a non-activating Arg
PT	residue - used in the treatment of haemophilias, by improvement of

PT haemostasis
XX Claim 30; Page -: 90pp; English.
CC AAM11330-W11472 represent active Factor VIII:C analogues of the
CC invention. These sequences were created by mutating the wild type Factor
CC VIII:C coding sequence (see AAT51337) using mutagenic primers. The
CC analogues comprise a native Factor VIII:C polypeptide modified at a site
CC adjacent to a non-activating Arg residue so that a Arg-Pro or Pro-Arg
CC dipeptide is created. Factor VIII:C is a large glycoprotein that
CC participates in the blood coagulation cascade that ultimately converts
CC soluble fibrinogen to insoluble fibrin clot, effecting haemostasis. A
CC deficiency in Factor VIII:C is responsible for haemophilia A, which is an
CC X-chromosome-linked inherited bleeding diathesis. Factor VIII:C is
CC activated by plasma proteases, such as thrombin. During activation the
CC mature polypeptide is cleaved to generate heavy and light chain fragments
CC that are further cleaved. Complexes of two or more of the analogues,
CC nucleic acids and vectors encoding them may be used alone or in
CC conjunction with each other, for the prevention or treatment of active
CC Factor VIII:C deficiency in a mammal. The analogues may be used as
CC immunogens to raise antibodies, and in the treatment of haemophilias, by
CC improvement of haemostasis. The analogues are resistant to proteolytic
CC cleavage and display increased plasma half-life. They may be administered
CC at lower dosages and by different modes of administration.
XX Sequence 2351 AA:
SQ
Query Match 100.0%; Score 12414; DB 18; Length 2351;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2350; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 MOTELSTCEFLCLARFCESATRRYYIGAVELSWDMQSDGLPVDARPPRPVSPFN 60
DB 1 MOTELSTCEFLCLARFCESATRRYYIGAVELSWDMQSDGLPVDARPPRPVSPFN 60
QY 61 TSVYVKKTLFVEETDHLFNIAKRPWMGLIPTIOAEVYTVVTLKNASHPSLAV 120
DB 61 TSVYVKKTLFVEETDHLFNIAKRPWMGLIPTIOAEVYTVVTLKNASHPSLAV 120
QY 121 GVSVMKASGAEYDQDSOREKEDDVFPFGSGHTVYMOVLKSNPMASDPLCLTSTYSH 180
DB 121 GVSVMKASGAEYDQDSOREKEDDVFPFGSGHTVYMOVLKSNPMASDPLCLTSTYSH 180
QY 121 GVSVMKASGAEYDQDSOREKEDDVFPFGSGHTVYMOVLKSNPMASDPLCLTSTYSH 180
DB 121 GVSVMKASGAEYDQDSOREKEDDVFPFGSGHTVYMOVLKSNPMASDPLCLTSTYSH 180
QY 181 VDLVNDLNSGLIGALVCRGSLAKKKTOTLKFTLLFAVDDESKSHSTKNSLMORD 240
DB 181 VDLVNDLNSGLIGALVCRGSLAKKKTOTLKFTLLFAVDDESKSHSTKNSLMORD 240
QY 241 AASARAMPKMTVNGVYVNSLPGLICGHRKSYVWVYIGMGTTPVHSIFLEGTELVNHH 300
DB 241 AASARAMPKMTVNGVYVNSLPGLICGHRKSYVWVYIGMGTTPVHSIFLEGTELVNHH 300
QY 301 RQASLEISPTFLAQTLLMDLQGLLFCHISSHQHDMEMAYVYVSCPEEPQLMKKNE 360
DB 301 RQASLEISPTFLAQTLLMDLQGLLFCHISSHQHDMEMAYVYVSCPEEPQLMKKNE 360
QY 361 EAEVDYDOLTDSEMDVYRFDDDSPSFIOIRSVAKKHKPTVWHYIAAEEDMDYAPVLA 420
DB 361 EAEVDYDOLTDSEMDVYRFDDDSPSFIOIRSVAKKHKPTVWHYIAAEEDMDYAPVLA 420
QY 421 PDDRSYKSOYLNGSPORIGRKYKRVFMAVYTDDEKTEALIOHESGLIPLLYGEVDTL 480
DB 421 PDDRSYKSOYLNGSPORIGRKYKRVFMAVYTDDEKTEALIOHESGLIPLLYGEVDTL 480
QY 481 LIIFKNOASRPNTYPIGTIDVRLYSRLPGVYHLKDPILLGELFFKTKMTVVEGCP 540
DB 481 LIIFKNOASRPNTYPIGTIDVRLYSRLPGVYHLKDPILLGELFFKTKMTVVEGCP 540
QY 541 TKSDFPCLTRYYSFVNMRDLASGLIGPLLICYKESVDQGNQIMSDKRNVIILSVDE 600
DB 541 TKSDFPCLTRYYSFVNMRDLASGLIGPLLICYKESVDQGNQIMSDKRNVIILSVDE 600
QY 601 NRSWYLTENIORFLPNPAGVQLEDPEFOASNTMHSINGYVDSIJQSYCHAEVAYWYLS 660
DB 601 NRSWYLTENIORFLPNPAGVQLEDPEFOASNTMHSINGYVDSIJQSYCHAEVAYWYLS 660

DB 601 NRSWYLTENIORFLPNPAGVQLEDPEFOASNTMHSINGYVDSIJQSYCHAEVAYWYLS 660
QY 661 IGAGQDFLSVFESGTYFKHKMYVEDLTLEFPSEGETFMSKENVGMLIIGCHNSDFNRG 720
DB 661 IGAGQDFLSVFESGTYFKHKMYVEDLTLEFPSEGETFMSKENVGMLIIGCHNSDFNRG 720
QY 721 MTRALKVSSCKNTGDIYEDSYDIAVYLSKNNALIPRPSQNSNHPSTROKQNNATTI 780
DB 721 MTRALKVSSCKNTGDIYEDSYDIAVYLSKNNALIPRPSQNSNHPSTROKQNNATTI 780
QY 781 PENDIEKTDPMFAHRTMPKIQNVSSDDLMLLRQSPTPGSLSDQAKETPSDDPS 840
DB 781 PENDIEKTDPMFAHRTMPKIQNVSSDDLMLLRQSPTPGSLSDQAKETPSDDPS 840
QY 841 PGALDSNNSLEMTNFRPOLHNSGDVFTPESSGQLRLNKKLTATATLKKLDPKYST 900
DB 841 PGALDSNNSLEMTNFRPOLHNSGDVFTPESSGQLRLNKKLTATATLKKLDPKYST 900
QY 901 SNNLTSTIPSDNLAAQTDNNTSLGPPSMVHYDSQDITLFGKSSPLTESGGLSSEE 960
DB 901 SNNLTSTIPSDNLAAQTDNNTSLGPPSMVHYDSQDITLFGKSSPLTESGGLSSEE 960
QY 961 NNDKILLESGLMNSQESSWCKNVSTESGRLFKGRAGPALLTKDNALFKVSIKLTN 1020
DB 961 NNDKILLESGLMNSQESSWCKNVSTESGRLFKGRAGPALLTKDNALFKVSIKLTN 1020
QY 1021 KTSNNSATRKTHIDGSLILENSPVMQNTLESPTFKKVTPLIHDMKMDKNTALRL 1080
DB 1021 KTSNNSATRKTHIDGSLILENSPVMQNTLESPTFKKVTPLIHDMKMDKNTALRL 1080
QY 1081 NMSKNTTSSKKNMVOQKKEGPIPPDAQNPDMSEFKMLPLPSANWIORTHGKNSLNG 1140
DB 1081 NMSKNTTSSKKNMVOQKKEGPIPPDAQNPDMSEFKMLPLPSANWIORTHGKNSLNG 1140
QY 1141 QGSPKQVLSGPEKSYEGONFLSEKNKVVVGKEFTKDGLEKEMTPSSRLFTNLND 1200
DB 1141 QGSPKQVLSGPEKSYEGONFLSEKNKVVVGKEFTKDGLEKEMTPSSRLFTNLND 1200
QY 1201 LHEHNTNHOEKKIOEIEKEKTELIONNVYLPDIHYVGTNPKNKLFLSTONVEGYD 1260
DB 1201 LHEHNTNHOEKKIOEIEKEKTELIONNVYLPDIHYVGTNPKNKLFLSTONVEGYD 1260
QY 1261 GAYAPYLDQFSLNDSTNRTKKTATAFHSKSGEENLGLQNOTKQIVERTACTTRISPT 1320
DB 1261 GAYAPYLDQFSLNDSTNRTKKTATAFHSKSGEENLGLQNOTKQIVERTACTTRISPT 1320
QY 1321 SOONFVOTRSKRALQFRLPEETELKERRIYVDTSTOWSKNNKHLPSTLQIDYNEKE 1380
DB 1321 SOONFVOTRSKRALQFRLPEETELKERRIYVDTSTOWSKNNKHLPSTLQIDYNEKE 1380
QY 1381 KGAITOSPISDCLFRSHSIPQANRSPPLTAAYVSPSTIRPIYLRVLPQDNSHLPASY 1440
DB 1381 KGAITOSPISDCLFRSHSIPQANRSPPLTAAYVSPSTIRPIYLRVLPQDNSHLPASY 1440
QY 1441 RKKDSGVQESSHFIQAGKAKNNLSLALILEMTGQORVGSIGTSANSTYKRYKENVLP 1500
DB 1441 RKKDSGVQESSHFIQAGKAKNNLSLALILEMTGQORVGSIGTSANSTYKRYKENVLP 1500
QY 1501 KPDLPTKSGKVELLPKHYIYOKDLFPTETSNBSGPHLDLVESGLQGTGALKNWEANRP 1560
DB 1501 KPDLPTKSGKVELLPKHYIYOKDLFPTETSNBSGPHLDLVESGLQGTGALKNWEANRP 1560
QY 1561 GVPPLRYATVETSSAKTSPSKLDPALMDNHYGQIPKPEKMSQKSEKPAFKKDTJLSL 1620
DB 1561 GVPPLRYATVETSSAKTSPSKLDPALMDNHYGQIPKPEKMSQKSEKPAFKKDTJLSL 1620
QY 1621 NACESNHAIAAINEGQKPEIEVYTAQKQRTERLCSQNPVYLRKHQREITRTTLQSDOE 1680
DB 1621 NACESNHAIAAINEGQKPEIEVYTAQKQRTERLCSQNPVYLRKHQREITRTTLQSDOE 1680
QY 1681 IDYDPTISEVKKKEDPDIYDEENOSPSFOKTRRHFFIAVYRLMDMGSSPHVYLR 1740
DB 1681 IDYDPTISEVKKKEDPDIYDEENOSPSFOKTRRHFFIAVYRLMDMGSSPHVYLR 1740

```

QY 1741 AOSGSVPQKRVFOEFTDSFTOPLYRGELENHGLGPYIRAEVEDINIMWTERNOASR 1800
DB 1741 AOSGSVPQKRVFOEFTDSFTOPLYRGELENHGLGPYIRAEVEDINIMWTERNOASR 1800
QY 1801 PYSTFSLISYEDROGAPRKATYKPKETKTFKVOHMAPTDEDECKANAYTSVY 1860
DB 1801 PYSTFSLISYEDROGAPRKATYKPKETKTFKVOHMAPTDEDECKANAYTSVY 1860
QY 1861 DLEKDVHSLGLPLVCHFTNLPARGROYVOEALFPTTIDETKSWTFENNERCRA 1920
DB 1861 DLEKDVHSLGLPLVCHFTNLPARGROYVOEALFPTTIDETKSWTFENNERCRA 1920
QY 1921 PCNIQMEDPTEKENVRFHAINGYIMDTLPGLVMAODORIMWYLLSGNSNENHSIHFSGH 1980
DB 1921 PCNIQMEDPTEKENVRFHAINGYIMDTLPGLVMAODORIMWYLLSGNSNENHSIHFSGH 1980
QY 1981 VFTVRKKEEYKALYNLYPGVFETVEMLPKSKATWREBCLIGEHLAGNSTLFLVYSNK 2040
DB 1981 VFTVRKKEEYKALYNLYPGVFETVEMLPKSKATWREBCLIGEHLAGNSTLFLVYSNK 2040
QY 2041 QTPLGMAAGHTRDPOIRASGOYGNAPKRLAHYSGSINAMSTKEPESTIKYDLAPMT 2100
DB 2041 QTPLGMAAGHTRDPOIRASGOYGNAPKRLAHYSGSINAMSTKEPESTIKYDLAPMT 2100
QY 2101 HGKKTGAKOKRSSLYISQFTIMYSLDGKKMOTTRKNSGTGLMVFEGNDSSGIKHNIFN 2160
DB 2101 HGKKTGAKOKRSSLYISQFTIMYSLDGKKMOTTRKNSGTGLMVFEGNDSSGIKHNIFN 2160
QY 2161 PPTIARYIRLAPTHYSIRSLRAMELNGCDLNSCSPMLMESKASIDAOITASSYETNFA 2220
DB 2161 PPTIARYIRLAPTHYSIRSLRAMELNGCDLNSCSPMLMESKASIDAOITASSYETNFA 2220
QY 2221 TMSPSKARLHLOGNSNMRPOVNNPKEMLOVDFOKTKMYGVTTQGVKSLTSMYKFEFL 2280
DB 2221 TMSPSKARLHLOGNSNMRPOVNNPKEMLOVDFOKTKMYGVTTQGVKSLTSMYKFEFL 2280
QY 2281 ISSSODGHWTLFQNGKRVKVFQGNDSPTPVNSLDPPLLTRLRILHFSQVWHDIALRM 2340
DB 2281 ISSSODGHWTLFQNGKRVKVFQGNDSPTPVNSLDPPLLTRLRILHFSQVWHDIALRM 2340
QY 2341 EYLGCEADOLY 2351
DB 2341 EYLGCEADOLY 2351

RESULT 7
AAW11415
ID AAW11416 standard; Protein: 2351 AA.
AC AAW11416;
XX 20-NOV-1997 (first entry)
DE Active Factor VIII:C analogue A774X.
XX
KW Factor VIII:C; analogue; glycoprotein; blood coagulation cascade;
KW fibrinogen; fibrin clot; haemostasis; haemophilia A; bleeding diathesis;
KW plasma protease; thrombin; immunogen; antibody; haemophilic; therapy;
KW proteolytic cleavage.
XX
OS Homo sapiens.
OS Synthetic.
XX
FH key Location/Qualifiers
FT Peptide 1..19
FT FT /note= "signal peptide"
FT Protein 20..2351
FT FT /note= "mature Factor VIII:C"
FT Region 20..1667
FT FT /note= "heavy chain fragment"
FT Modified-site 774
FT /label= Phe, Glu, Pro

```

```

FT Region 1668..2350
FT FT /note= "light chain fragment"
FT FT Domain 760..1667
FT FT /note= "B domain"
XX MO9703195-A1.
XX
XX 30-JAN-1997.
XX
XX 09-JUL-1996; 96WD-US11444.
XX
XX 11-JUL-1995; 95US-0001025.
XX
XX (CHIR ) CHIRON CORP.
XX
XX Cohen FE, Hung DF, Innis M;
XX
XX WPI: 1997-119050/11.
XX
XX Factor VIII:C analog modified adjacent to a non-activating Arg
XX residue - used in the treatment of haemophilias, by improvement of
XX haemostasis
XX
XX Claim 26; Page -: 90pp; English.
XX
XX AAW11330-W1472 represent active Factor VIII:C analogues of the
XX invention. These sequences were created by mutating the wild type Factor
XX VIII:C coding sequence (see AAW51357) using mutagenic primers. The
XX analogues comprise a native Factor VIII:C polypeptide modified at a site
XX adjacent to a non-activating Arg residue so that a Arg-Pro or Pro-Arg
XX dipeptide is created. Factor VIII:C is a large glycoprotein that
XX participates in the blood coagulation cascade that ultimately converts
XX soluble fibrinogen to insoluble fibrin clot, effecting haemostasis. A
XX deficiency in Factor VIII:C is responsible for haemophilia A, which is an
XX X-chromosome-linked inherited bleeding diathesis. Factor VIII:C is
XX activated by plasma proteases, such as thrombin. During activation the
XX mature polypeptide is cleaved to generate heavy and light chain fragments
XX that are further cleaved. Complexes of two or more of the analogues,
XX nucleic acids and vectors encoding them may be used alone or in
XX conjunction with each other, for the prevention or treatment of active
XX Factor VIII:C deficiency in a mammal. The analogues may be used as
XX immunogens to raise antibodies, and in the treatment of haemophilias, by
XX cleavage and display increased plasma half-life. They may be administered
XX at lower dosages and by different modes of administration.
XX
XX Sequence 2351 AA:
XX
XX Query Match 100.0%; Score 12414; DB 18; Length 2351;
XX Best Local Similarity 100.0%; Pred. No. 0;
XX Matches 2350; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 MOELSTCFPLCLLRPCFSATRRYYLGAVELSDVYQSDGLDELVDARFPFPRVPSFPFN 60
DB 1 MOELSTCFPLCLLRPCFSATRRYYLGAVELSDVYQSDGLDELVDARFPFPRVPSFPFN 60
QY 61 TSYYKKTLVEFTDLFLFIKAPRPMMGLGTIOAEYVDVYITLKNAASPVSLAHV 120
DB 61 TSYYKKTLVEFTDLFLFIKAPRPMMGLGTIOAEYVDVYITLKNAASPVSLAHV 120
QY 121 GVSYTKASBEAEYDQTSOREKEDKTPFGSGTYYWQYTKENKPMASDPLCTYSYSH 180
DB 121 GVSYTKASBEAEYDQTSOREKEDKTPFGSGTYYWQYTKENKPMASDPLCTYSYSH 180
QY 181 VDLVKDNLNSGLIGALLVCRBSLAKETQTLKFTLLRAYPDEGKSHSETKSLMQDRD 240
DB 181 VDLVKDNLNSGLIGALLVCRBSLAKETQTLKFTLLRAYPDEGKSHSETKSLMQDRD 240
QY 241 AASARAPKMHVNGVYNSLPGILGCHRSYVMHWIGMGTTPVHSLFEGHTFLVRRH 300
DB 241 AASARAPKMHVNGVYNSLPGILGCHRSYVMHWIGMGTTPVHSLFEGHTFLVRRH 300
QY 301 ROASLEISPTTFLTAQTLMDLGOFLFCHISSHQHDSMEAYVAVDSCPEEPQLRMKNE 360

```


|||||
301 KQASLEISPTIFLQFLADLQFLFCCHSSQHDGMEATYKVDSCPEEPQJMKMNE 360
QY 361 EAEDYDDDLTDSEMDVYRFDDDNSSPFIQIRSAKKHPRKWHYIAABEDMDYAPLVTA 420
Db 361 EAEYDDDLTDSEMDVYRFDDDNSSPFIQIRSAKKHPRKWHYIAABEDMDYAPLVTA 420
QY 421 PDRSRYQSYLNGNGPQIRGKRYKKRYFMAVYDETFTKREAIQHESGILGPIILYGEVDTL 480
Db 421 PDRSRYQSYLNGNGPQIRGKRYKKRYFMAVYDETFTKREAIQHESGILGPIILYGEVDTL 480
QY 481 LTIIFKNQASRPYNYTPHGITDVRPLYSRRLPKGKHLKDFILPEITPKYKWTYEDGP 540
Db 481 LTIIFKNQASRPYNYTPHGITDVRPLYSRRLPKGKHLKDFILPEITPKYKWTYEDGP 540
QY 541 TKSDFRCILTRYSSFYVMERDLASGLIGPLILCYKESVQQRQNMISKRNVILFSYFDE 600
Db 541 TKSDFRCILTRYSSFYVMERDLASGLIGPLILCYKESVQQRQNMISKRNVILFSYFDE 600
QY 601 NRSWYLTENIQRLPNPAGVLEDEPFQASNMHSINGVYFSDLSVCLHEVAYWYLLS 660
Db 601 NRSWYLTENIQRLPNPAGVLEDEPFQASNMHSINGVYFSDLSVCLHEVAYWYLLS 660
QY 661 IGAQTDPLSYFSGYTFEKKHMYEDTLTFPFSGETVFMKMPGIMTIGCHNDFENRG 720
Db 661 IGAQTDPLSYFSGYTFEKKHMYEDTLTFPFSGETVFMKMPGIMTIGCHNDFENRG 720
QY 721 MTALKVSSCDKNTGOVYEDSYEDISAVILSKNNAIEPRFSQNSRHSSTQKQFNATYI 780
Db 721 MTALKVSSCDKNTGOVYEDSYEDISAVILSKNNAIEPRFSQNSRHSSTQKQFNATYI 780
QY 781 PENDIEKTDWFAHRPMKPIQNVSSSDMLMLNOSPPIGSLSDLAEEVYFSDPS 840
Db 781 PENDIEKTDWFAHRPMKPIQNVSSSDMLMLNOSPPIGSLSDLAEEVYFSDPS 840
QY 841 PGATDSNNSLSEMTHPROLHSGDMVTPPSGIGQLFLNEKIGTTNABELKKDIFYKST 900
Db 841 PGATDSNNSLSEMTHPROLHSGDMVTPPSGIGQLFLNEKIGTTNABELKKDIFYKST 900
QY 901 SNMLISTIPSDNLAAGTDNTSLGAPSPVHYDSQLDITLFGKSSPILTESGGLSLSEE 960
Db 901 SNMLISTIPSDNLAAGTDNTSLGAPSPVHYDSQLDITLFGKSSPILTESGGLSLSEE 960
QY 961 NNDKILSEGLMNSQESMGKNVSTESGRLFKGRRAHGPAULTKONMLFYVSLIKTN 1020
Db 961 NNDKILSEGLMNSQESMGKNVSTESGRLFKGRRAHGPAULTKONMLFYVSLIKTN 1020
QY 1021 KTSNNSATNRKTHDGSLLIENSVMQNTLSDTEFKKVTPLJHDMKMDKNAATLRL 1080
Db 1021 KTSNNSATNRKTHDGSLLIENSVMQNTLSDTEFKKVTPLJHDMKMDKNAATLRL 1080
QY 1081 NMSNKTTSNKNMEMVOQKKEGPIPPDQNPDMSPFFMLFLPESARWIORTHGKNSLNSG 1140
Db 1081 NMSNKTTSNKNMEMVOQKKEGPIPPDQNPDMSPFFMLFLPESARWIORTHGKNSLNSG 1140
QY 1141 OGSFPRKQVLSGPEKSEYQGNFLSEKNVYVVGKEFTKDGKLEKEMFSSRRLFLTLMDN 1200
Db 1141 OGSFPRKQVLSGPEKSEYQGNFLSEKNVYVVGKEFTKDGKLEKEMFSSRRLFLTLMDN 1200
QY 1201 LHENTHNQEKKIOEEIEKKEKTLIOENVYLPQIHVYGTGNFKNLFLLSTRONVBSYD 1260
Db 1201 LHENTHNQEKKIOEEIEKKEKTLIOENVYLPQIHVYGTGNFKNLFLLSTRONVBSYD 1260
QY 1261 GAYAPYLODFRSLNDSTNRKTHAFSKGEBENLGLONQKQIVERYACTRIRISPT 1320
Db 1261 GAYAPYLODFRSLNDSTNRKTHAFSKGEBENLGLONQKQIVERYACTRIRISPT 1320
QY 1321 SOONFVTOBSKRALQKQRLPLEETELERKIIVDDTSTQMSKNNKHLTPSTLLOIDYNEKE 1380
Db 1321 SOONFVTOBSKRALQKQRLPLEETELERKIIVDDTSTQMSKNNKHLTPSTLLOIDYNEKE 1380
QY 1381 KGATOSPISDCLTRSHSIPQANRSPPLIAKVSFSPSIRIYITRVLFDONSSHLPASY 1440
Db 1381 KGATOSPISDCLTRSHSIPQANRSPPLIAKVSFSPSIRIYITRVLFDONSSHLPASY 1440

Db 1381 KGATOSPISDCLTRSHSIPQANRSPPLIAKVSFSPSIRIYITRVLFDONSSHLPASY 1440
QY 1441 RKQSGVOESSHFLQGAKKNNLSLALTLLEMGDQREYGLGTSAINSTYKRYNLYP 1500
Db 1441 RKQSGVOESSHFLQGAKKNNLSLALTLLEMGDQREYGLGTSAINSTYKRYNLYP 1500
QY 1501 KPDLPTSGKVELLPKVHIOKDLFLETSSNGSPGHLDLVGLSLGTEGAIKRMEANRP 1560
Db 1501 KPDLPTSGKVELLPKVHIOKDLFLETSSNGSPGHLDLVGLSLGTEGAIKRMEANRP 1560
QY 1561 GRVPLRVATESAKTPSKLLDPLAMDNYGQIQIEBEMKSOEKSEPKTAFAKKDPTLSTL 1620
Db 1561 GRVPLRVATESAKTPSKLLDPLAMDNYGQIQIEBEMKSOEKSEPKTAFAKKDPTLSTL 1620
QY 1621 NAGESNHATAINEQONKPELEVTAKQORTERLCSQNPPLAKRHOREITRTLOSDEE 1680
Db 1621 NAGESNHATAINEQONKPELEVTAKQORTERLCSQNPPLAKRHOREITRTLOSDEE 1680
QY 1681 IDYDDTISYEMKKEFDIYDEDEQSPRSFQKTRHYFLAVERLMDYGASSPVLNKR 1740
Db 1681 IDYDDTISYEMKKEFDIYDEDEQSPRSFQKTRHYFLAVERLMDYGASSPVLNKR 1740
QY 1741 AQSQSVPOKRYVQEFITDGSFTQPLRGELNEHLGLGPYIRAEVEDNIMYTFPNQASR 1800
Db 1741 AQSQSVPOKRYVQEFITDGSFTQPLRGELNEHLGLGPYIRAEVEDNIMYTFPNQASR 1800
QY 1801 PYSFYSLSLSEEDQROGAEBPRKNFYVPMETKYTFMKYQHMAFTKDEFCKAMAFYFDV 1860
Db 1801 PYSFYSLSLSEEDQROGAEBPRKNFYVPMETKYTFMKYQHMAFTKDEFCKAMAFYFDV 1860
QY 1861 DLEKDYHSGILGPLLYCHTNTLANAGROYTVOEFALFFITFDETSWYTFENMRNRA 1920
Db 1861 DLEKDYHSGILGPLLYCHTNTLANAGROYTVOEFALFFITFDETSWYTFENMRNRA 1920
QY 1921 PCNTQMEDPFEKENTREFAINGTIMOTLACIYMAODORIPRYLLSMQSNENHSHIESGH 1980
Db 1921 PCNTQMEDPFEKENTREFAINGTIMOTLACIYMAODORIPRYLLSMQSNENHSHIESGH 1980
QY 1981 VFTVRRKKEYNALNLYPGVETEMLPKAGIWRVCLLIGELHLAGNSTLPLVYSNK 2040
Db 1981 VFTVRRKKEYNALNLYPGVETEMLPKAGIWRVCLLIGELHLAGNSTLPLVYSNK 2040
QY 2041 QTPLGMAHSHIDPQITASGOYGOVAPKRLARLHYSQGSINANSTKEPSWIKYDLPAMI 2100
Db 2041 QTPLGMAHSHIDPQITASGOYGOVAPKRLARLHYSQGSINANSTKEPSWIKYDLPAMI 2100
QY 2101 HGITQGAOROKFSSLYISQFTIMYSLDGKKMOTYRGSNGTLMVFGNDSSGIRKHNFN 2160
Db 2101 HGITQGAOROKFSSLYISQFTIMYSLDGKKMOTYRGSNGTLMVFGNDSSGIRKHNFN 2160
QY 2161 PPIIARIYRLHPHYSISTYRMLMCCDLNCSMPKLNESKALISDAQITASSYTFNMA 2220
Db 2161 PPIIARIYRLHPHYSISTYRMLMCCDLNCSMPKLNESKALISDAQITASSYTFNMA 2220
QY 2221 TWSPSKARLHLQGRSNANRPQVNNPKEMLOVDFOTMRYVTGTTQGVKSLTSMYKFE 2280
Db 2221 TWSPSKARLHLQGRSNANRPQVNNPKEMLOVDFOTMRYVTGTTQGVKSLTSMYKFE 2280
QY 2281 ISSSODQHWTLEFONGKAVKYFQGNQDSFPVNVSLDPLLTRILRIHPQSWHQALARM 2340
Db 2281 ISSSODQHWTLEFONGKAVKYFQGNQDSFPVNVSLDPLLTRILRIHPQSWHQALARM 2340
QY 2341 EYLGCEADQLY 2351
Db 2341 EYLGCEADQLY 2351
RESULT 8
AAW11343
ID AAW11343 standard; Protein; 2351 AA.
XX
AC AAW11343;
XX


```

Db      1081  |||||
Oy      1141  OGPSPKOLVSLGPERSVEGONFLSEKNKVVYVKGEEFTKDVLKEMVPSRRILFTLNIDN 1200
Db      1141  OGPSPKOLVSLGPERSVEGONFLSEKNKVVYVKGEEFTKDVLKEMVPSRRILFTLNIDN 1200
Oy      1201  LHENNTHNOEKKIOEELKEKTELLIOENVYLPQIHVYVGTNPKFNKLFILSTRQNVESYD 1260
Db      1201  LHENNTHNOEKKIOEELKEKTELLIOENVYLPQIHVYVGTNPKFNKLFILSTRQNVESYD 1260
Oy      1261  GAYAPYLODFRSLNDSTNRTKKTAFNSKKEEENLGLGNQTKQIYEKYACTTRISPT 1320
Db      1261  GAYAPYLODFRSLNDSTNRTKKTAFNSKKEEENLGLGNQTKQIYEKYACTTRISPT 1320
Oy      1321  SOONFVOTORSKRALKOPRLPLEETLEKRIIYDSTQWSKNNKHLTPSTLQIDYNEKE 1380
Db      1321  SOONFVOTORSKRALKOPRLPLEETLEKRIIYDSTQWSKNNKHLTPSTLQIDYNEKE 1380
Oy      1381  KGAIQSPSLDCLTRSHSIPQANRSPPLIAKYSSPSPRIYLRVLPQDNSSHTLPASY 1440
Db      1381  KGAIQSPSLDCLTRSHSIPQANRSPPLIAKYSSPSPRIYLRVLPQDNSSHTLPASY 1440
Oy      1441  RKKGSGVOESHTLOGAKNNLSLALILEMTGDQREVSLGTSAVNSVYKRVENVLP 1500
Db      1441  RKKGSGVOESHTLOGAKNNLSLALILEMTGDQREVSLGTSAVNSVYKRVENVLP 1500
Oy      1501  KPDLPTSGKYELLPKVHIYQKDLPEFTETSNQSPGHLDIVESLLOGTGALKMEANRP 1560
Db      1501  KPDLPTSGKYELLPKVHIYQKDLPEFTETSNQSPGHLDIVESLLOGTGALKMEANRP 1560
Oy      1561  GAVPLRVATESAKPSPSKLIDPLANDNHYGOLPKEMKSOEKSEPKAFKKDTILSL 1620
Db      1561  GAVPLRVATESAKPSPSKLIDPLANDNHYGOLPKEMKSOEKSEPKAFKKDTILSL 1620
Oy      1621  NACESNHAIAINNGOKKPEIETVNAKOGREBRLCSQNPVYLKRHOREITRTTLOSDEE 1680
Db      1621  NACESNHAIAINNGOKKPEIETVNAKOGREBRLCSQNPVYLKRHOREITRTTLOSDEE 1680
Oy      1681  IDYDPTISENKKEDFDIYDEENOSPRSFOKTRHYFAIVARLMDYGMSSSPHYLNR 1740
Db      1681  IDYDPTISENKKEDFDIYDEENOSPRSFOKTRHYFAIVARLMDYGMSSSPHYLNR 1740
Oy      1741  AAGSGVPOFKKVFQEEPTDGSFTQPLRYGELNEHGLGPIYRAVEDNIMVTFENQASR 1800
Db      1741  AAGSGVPOFKKVFQEEPTDGSFTQPLRYGELNEHGLGPIYRAVEDNIMVTFENQASR 1800
Oy      1801  PYSFYSLSIYEEOROGAPRKPNFYKPNKTYTVMKYQHHMAPTYDEEDCKAMAYESDV 1860
Db      1801  PYSFYSLSIYEEOROGAPRKPNFYKPNKTYTVMKYQHHMAPTYDEEDCKAMAYESDV 1860
Oy      1861  DLEKDVHSGLIGPLLVCHTNTLNPAGROYVQEFALFTIFDETRKSWYFENMERNCRA 1920
Db      1861  DLEKDVHSGLIGPLLVCHTNTLNPAGROYVQEFALFTIFDETRKSWYFENMERNCRA 1920
Oy      1921  PCNIOMEDPTFKENYRHAINGYIMDTLPGLVMAQDORIMVYILSMGSNENHSHFEGH 1980
Db      1921  PCNIOMEDPTFKENYRHAINGYIMDTLPGLVMAQDORIMVYILSMGSNENHSHFEGH 1980
Oy      1981  VFTVVRKKEEYKALYNYPGVFETVEMLPKRAGINRVBCLIGHLHAGSTLFLVYSNKC 2040
Db      1981  VFTVVRKKEEYKALYNYPGVFETVEMLPKRAGINRVBCLIGHLHAGSTLFLVYSNKC 2040
Oy      2041  QTPGLMASGHIRDFQITASGOYGMAPKLARLYSGSINAMSTKEPFSWIKDOLAPMI 2100
Db      2041  QTPGLMASGHIRDFQITASGOYGMAPKLARLYSGSINAMSTKEPFSWIKDOLAPMI 2100
Oy      2101  HGKIQGAKOKFSSYISOTILMSDGKKWTTIGNSSTGLLWFTFGVNDSSGILHNFTN 2160
Db      2161  PPIIARYLRAPHTYSIRSTLRNMLMGCDLNSCMLGMSKRAISDAQTASSYFTNMFA 2220

```

```

Db      2161  PPIIARYLRAPHTYSIRSTLRNMLMGCDLNSCMLGMSKRAISDAQTASSYFTNMFA 2220
Oy      2221  TWSPSKARLHLQGRSNAMRPQVNNPREMLQVDFQTKAKYVGYTQGVNSLLTSMYVEFL 2280
Db      2221  TWSPSKARLHLQGRSNAMRPQVNNPREMLQVDFQTKAKYVGYTQGVNSLLTSMYVEFL 2280
Oy      2281  ISSSODGHQWTLFPQNGKRYKVFQGNDSFTPVVNSLDPPLLPRLYRIRHPOSMVHOIALRM 2340
Db      2281  ISSSODGHQWTLFPQNGKRYKVFQGNDSFTPVVNSLDPPLLPRLYRIRHPOSMVHOIALRM 2340
Oy      2341  EYLGEADQLX 2351
Db      2341  EYLGEADQLX 2351

RESULT 9
AAW11461
ID      AAW11461 standard; Protein; 2351 AA.
XX
AC      AAW11461;
XX
DT      20-NOV-1997 (first entry)
XX
DE      Active Factor VIII:C analogue V1717X.
XX
KW      Factor VIII:C; analogue; glycoprotein; blood coagulation cascade;
KW      fibrinogen; fibrin clot; haemostasis; haemophilia A; bleeding diathesis;
KW      plasma protease; thrombin; immunogen; antibody; haemophilic therapy;
KW      proteolytic cleavage.
XX
OS      Homo sapiens.
XX
XX      Synthetic.
XX
FH      Key
FH      Location/Qualifiers
FH      Peptide
FH      1..19
FH      /note="signal peptide"
FH      20..2351
FH      Protein
FH      /note="mature Factor VIII:C"
FH      Region
FH      20..1667
FH      /note="heavy chain fragment"
FH      Region
FH      1668..2350
FH      /note="light chain fragment"
FH      760..1667
FH      Domain
FH      /note="B domain"
FH      Modified-site
FH      1736
FH      /label="Phe, Glu, Pro

W09703195-A1.
PD      30-JAN-1997.
PF      09-JUL-1996; 96WO-US11444.
PR      11-JUL-1995; 95US-0001025.
PA      (CHIR ) CHIRON CORP.
PI      Cohen FE, Hung DT, Innis M;
PI      WPI; 1997-119050/11.
CC      Factor VIII:C analog modified adjacent to a non-activating Arg
CC      residue - used in the treatment of haemophiliacs, by improvement of
CC      haemostasis
CC      Claim 36; Page -: 90pp; English.
CC      AAW11330-W11472 represent active Factor VIII:C analogues of the
CC      invention. These sequences were created by mutating the wild type Factor
CC      VIII:C coding sequence (see AAV15157) using mutagenic primers. The
CC      analogues comprise a native Factor VIII:C polypeptide modified at a site
CC      adjacent to a non-activating Arg residue so that a Arg-Pro or Pro-Arg
CC      dipeptide is created. Factor VIII:C is a large glycoprotein that

```

CC participate in the blood coagulation cascade that ultimately converts
CC soluble fibrinogen to insoluble fibrin clot, effecting haemostasis. A
CC deficiency in Factor VIII:C is responsible for haemophilia A, which is an
CC X-chromosome-linked inherited bleeding diathesis. Factor VIII:C is
CC activated by plasma proteases, such as thrombin. During activation the
CC mature polypeptide is cleaved to generate heavy and light chain fragments
CC that are further cleaved. Complexes of two or more of the analogues,
CC nucleic acids and vectors encoding them may be used alone or in
CC conjunction with each other, for the prevention or treatment of active
CC Factor VIII:C deficiency in a mammal. The analogues may be used as
CC immunogens to raise antibodies, and in the treatment of hemophilias, by
CC improvement of haemostasis. The analogues are resistant to proteolytic
CC cleavage and display increased plasma half-life. They may be administered
CC at lower dosages and by different modes of administration.

SQ Sequence 2351 AA;

Query Match	100.0%;	Score 12413;	DB 18;	Length 2351;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 2350; Conservative	0;	Mismatches	1;	Indels 0; Gaps 0;

QY	1	MOJELSTCEPCLLJPCFESATRRYLAGVELSMIDMYQMOJLELVDADRPPVPSPEN	60
Db	1	MOJELSTCEPCLLJPCFESATRRYLAGVELSMIDMYQMOJLELVDADRPPVPSPEN	60
QY	61	TSVYVYKTLFPEFBDHLFJNAPRPPMGLGPTIOAEVYTYVTYITLKMAHSPLAHV	120
Db	61	TSVYVYKTLFPEFBDHLFJNAPRPPMGLGPTIOAEVYTYVTYITLKMAHSPLAHV	120
QY	121	GVSYVYKASGAEYDQOTSOREKEDDKVPGGSHYVYVWLKENGPMASDPLCLVSYLSH	180
Db	121	GVSYVYKASGAEYDQOTSOREKEDDKVPGGSHYVYVWLKENGPMASDPLCLVSYLSH	180
QY	181	VDVLYDNLSGLIGALVLCRGSGSLAKEKOTJHFFILFPAVYDEKSWHSSTKNSLMODRD	240
Db	181	VDVLYDNLSGLIGALVLCRGSGSLAKEKOTJHFFILFPAVYDEKSWHSSTKNSLMODRD	240
QY	241	AASARAPMHTVNGVYNSRLPGLICGRKSYVMYHIGAGTTPVHSHIFLEGHFLVYNH	300
Db	241	AASARAPMHTVNGVYNSRLPGLICGRKSYVMYHIGAGTTPVHSHIFLEGHFLVYNH	300
QY	301	ROSLEISPTIFLPAOTLMDJGOFILFCHSHSHOHGMEAVYVDSCPREPOLMKRNE	360
Db	301	ROSLEISPTIFLPAOTLMDJGOFILFCHSHSHOHGMEAVYVDSCPREPOLMKRNE	360
QY	361	EADYDQDLOTSEMDVYRPDDNSPSTQIRSAFKHPTWYHIAAESEMDYAPVLA	420
Db	361	EADYDQDLOTSEMDVYRPDDNSPSTQIRSAFKHPTWYHIAAESEMDYAPVLA	420
QY	421	PDRSTKSOYLYNNGPOTIGKTIKAVYRPAAYDETFTFTRALIOHESGLIPLLYGEVGDTL	480
Db	421	PDRSTKSOYLYNNGPOTIGKTIKAVYRPAAYDETFTFTRALIOHESGLIPLLYGEVGDTL	480
QY	481	LILIFKNQASRPANYTPHGLIDVAPLYSRRLPKGVKMLDFFLIGELIFKKTWYVEGCP	540
Db	481	LILIFKNQASRPANYTPHGLIDVAPLYSRRLPKGVKMLDFFLIGELIFKKTWYVEGCP	540
QY	541	TKSDPCLLKRYYSSFFVMMERDLASGLIPLLYICFESVYOQRONJMSDKANVILFSYDE	600
Db	541	TKSDPCLLKRYYSSFFVMMERDLASGLIPLLYICFESVYOQRONJMSDKANVILFSYDE	600
QY	601	NBSWYTLLENIORPLPNAGQOLEDEPAPASIMHISINGVDSIQJSLVCLHAEVAYWILS	660
Db	601	NBSWYTLLENIORPLPNAGQOLEDEPAPASIMHISINGVDSIQJSLVCLHAEVAYWILS	660
QY	661	IGADOTPLSVFPGSGTFRKKNVYEDTLTLTFPGSGEFTVPMSENGMLITLCHNSVDFRNG	720
Db	661	IGADOTPLSVFPGSGTFRKKNVYEDTLTLTFPGSGEFTVPMSENGMLITLCHNSVDFRNG	720
QY	721	MTALLVYSSCOKNMGDYEDJESYEDISALVLSKNNALTEPSEFSOHSRBPSTROKQENATTI	780
Db	721	MTALLVYSSCOKNMGDYEDJESYEDISALVLSKNNALTEPSEFSOHSRBPSTROKQENATTI	780

QY	781	IBENDIERTDWMFNRHTRPMPKIQWVSSDDLMLRLRSPPHGLSLSDLOQAKKETFSDPS	840
Db	781	PENDIERTDWMFNRHTRPMPKIQWVSSDDLMLRLRSPPHGLSLSDLOQAKKETFSDPS	840
QY	841	PGALDSNNLSLEMTFRPDLHSGDMVFVPEBSGLRLNKEKLTGAATLAKLDFKVSST	900
Db	841	PGALDSNNLSLEMTFRPDLHSGDMVFVPEBSGLRLNKEKLTGAATLAKLDFKVSST	900
QY	901	SNMLISTIPBDNLAAGTNTSSLAGPPMPVPHVDSOLDLTTLGGKSSPLESGPJSLSSE	960
Db	901	SNMLISTIPBDNLAAGTNTSSLAGPPMPVPHVDSOLDLTTLGGKSSPLESGPJSLSSE	960
QY	961	NNDSKLLSEGLNNSOESSMGKNVSTSSGGLFPGKRAIGPALLKONALFKYSISLKTN	1020
Db	961	NNDSKLLSEGLNNSOESSMGKNVSTSSGGLFPGKRAIGPALLKONALFKYSISLKTN	1020
QY	1021	KTSNNKSNTRKTHIDOPSLLENSPSWQNLLESTOEKRTPLIHDRMLDNKNATLARL	1080
Db	1021	KTSNNKSNTRKTHIDOPSLLENSPSWQNLLESTOEKRTPLIHDRMLDNKNATLARL	1080
QY	1081	NHMSKNTTSSKNEMVQOKKGGPIPDQONPDMKSPFKMLIPESAMVIOPTHGKNSLNG	1140
Db	1081	NHMSKNTTSSKNEMVQOKKGGPIPDQONPDMKSPFKMLIPESAMVIOPTHGKNSLNG	1140
QY	1141	QGPSPKOLVJLGEKSEVEGONFLSEKKNVVGKGEFTDVLKEMVPSRRMLFTNLND	1200
Db	1141	QGPSPKOLVJLGEKSEVEGONFLSEKKNVVGKGEFTDVLKEMVPSRRMLFTNLND	1200
QY	1201	LHENNTNOKKIOEIEIEKKEKTLIOENVVLPOLHVTGKTKPMNLFILSTONVSGYD	1260
Db	1201	LHENNTNOKKIOEIEIEKKEKTLIOENVVLPOLHVTGKTKPMNLFILSTONVSGYD	1260
QY	1261	GAYAPVJQDPRSLNDSTNRKTKKTAHNSKKGEBENLEGLQNOTQIYEKACTYISNT	1320
Db	1261	GAYAPVJQDPRSLNDSTNRKTKKTAHNSKKGEBENLEGLQNOTQIYEKACTYISNT	1320
QY	1321	SOONFVTOQRKRALKQFRLPLEETLEKRLIYVDSTOMSKNMKHLPTLQIDYNEKE	1380
Db	1321	SOONFVTOQRKRALKQFRLPLEETLEKRLIYVDSTOMSKNMKHLPTLQIDYNEKE	1380
QY	1381	KGALITOSPDLDCITRSHSIPQANRSPRIAKVSSPFSIRPIYTLRVLFEQONSHDPAASY	1440
Db	1381	KGALITOSPDLDCITRSHSIPQANRSPRIAKVSSPFSIRPIYTLRVLFEQONSHDPAASY	1440
QY	1441	RKDGVOESHFLOQAKKNLSLAILTEMTDQOEGVSGTATNSVYKKEVNTVLP	1500
Db	1441	RKDGVOESHFLOQAKKNLSLAILTEMTDQOEGVSGTATNSVYKKEVNTVLP	1500
QY	1501	RKDLPTKSGVVELPVNHIYOKDLFPHETNSPGHLDVBSGLQGTBOALIKKNENARP	1560
Db	1501	RKDLPTKSGVVELPVNHIYOKDLFPHETNSPGHLDVBSGLQGTBOALIKKNENARP	1560
QY	1561	GVPPFLRVATESSAKTPSKLDPJLAMDNIHGTQIPEEKKSQEKSPKTAFAKKDTIISL	1620
Db	1561	GVPPFLRVATESSAKTPSKLDPJLAMDNIHGTQIPEEKKSQEKSPKTAFAKKDTIISL	1620
QY	1621	NACSENNAIAINEGONKRELEVYMAKOGTETLCSONPVJLKHQOELIETRTLQSDOE	1680
Db	1621	NACSENNAIAINEGONKRELEVYMAKOGTETLCSONPVJLKHQOELIETRTLQSDOE	1680
QY	1681	IDYDITISVEKKEDDIIYDEDENOSPQKTRHYFAAVERLAMPYUGSSPVLNNR	1740
Db	1681	IDYDITISVEKKEDDIIYDEDENOSPQKTRHYFAAVERLAMPYUGSSPVLNNR	1740
QY	1741	AOSGSVPOFKVVFQEFJTGSGFTQPLRGLNHNHJLGLGPTIYAEVDNIAMTFRQNSR	1800
Db	1741	AOSGSVPOFKVVFQEFJTGSGFTQPLRGLNHNHJLGLGPTIYAEVDNIAMTFRQNSR	1800
QY	1801	PVSFVSSLISIEEDROGAEPKPNVAPNETKTYFPMKVQHNMAPTKOEFCRANAYFSDV	1860
Db	1801	PVSFVSSLISIEEDROGAEPKPNVAPNETKTYFPMKVQHNMAPTKOEFCRANAYFSDV	1860
QY	1861	DLEKVVHSGILGILVACHNTNLPNABORQTVQDEFALFETTFEDTBSGMYETEMBRNCRA	1920

Dd		1861	DLEKDVHSGILGICLVCCHNTLNINPAHQRYVQEFALFPTTJEDETKMVFTEENNERNCIA	1920
Oy		1921	PCNIDMDEPFKEKNRFRPHNGSYIMDTLPGLVMAODORIRWYLLSMGSNNENHSIHFSGH	1980
Dd		1921	PCNIQMDEPFKEKNRFRPHNGSYIMDTLPGLVMAODORIRWYLLSMGSNNENHSIHFSGH	1980
Oy		1981	VFTYRKKEEYKNAIYLIPGVFEIVEMLSBKAGITMVBCLLGEHLHAGNSTPLVYSNRK	2040
Dd		1981	VFTYRKKEEYKNAIYLIPGVFEIVEMLSBKAGITMVBCLLGEHLHAGNSTPLVYSNRK	2040
Oy		2041	QTPLGMASGHIRDFOITASGOYGWAPEKLARLHYSGISINAMSTYEPPSWIKVDLLAPMI	2100
Dd		2041	QTPLGMASGHIRDFOITASGOYGWAPEKLARLHYSGISINAMSTYEPPSWIKVDLLAPMI	2100
Oy		2101	HGITQOARKEFSLSLYISOFLIMYSLLDGKRMQTYRNNSGTGLMFEGNVDSOGIKHNIN	2160
Dd		2101	HGITQOARKEFSLSLYISOFLIMYSLLDGKRMQTYRNNSGTGLMFEGNVDSOGIKHNIN	2160
Oy		2161	PPIIARIYRIIHPHYISINSETLMELMGCDLNSCSMPLGKESKATSDAQITASSYFNMPA	2220
Dd		2161	PPIIARIYRIIHPHYISINSETLMELMGCDLNSCSMPLGKESKATSDAQITASSYFNMPA	2220
Oy		2221	TWPSKARLILOGRSNAMPPOYNPKREMLQVDRCKMYKTGYTTOGYKSILTSMYKEPL	2280
Dd		2221	TWPSKARLILOGRSNAMPPOYNPKREMLQVDRCKMYKTGYTTOGYKSILTSMYKEPL	2280
Oy		2281	ISSSQDHQHTLFEPONGKYKVPQGNDSPFPVYNSLDPELTRIRLRHPOSWYHOIALNM	2340
Dd		2281	ISSSQDHQHTLFEPONGKYKVPQGNDSPFPVYNSLDPELTRIRLRHPOSWYHOIALNM	2340
Oy		2341	EVLGCEADQLY 2351 	
Dd		2341	EVLGCEADQLY 2351 	
RESULT 10				
AA	11	44	5	
XX	ID	AA	114445 standard; Protein; 2351 AA.	
AC				
XX		AA	114445;	
DT				
XX		20-NOV-1997	(first entry)	
DE				
XX		Active Factor VIII:C analogue L1643X.		
KW		Factor VIII:C; analogue; glycoprotein; blood coagulation cascade;		
KM		fibrinogen; fibrin clot; haemostasis; haemophilia A; bleeding diathesis;		
RW		plasma protease; thrombin; immunogen; antibody; haemophilic; therapy;		
KW		proteolytic cleavage.		
XX				
OS		Homo sapiens.		
OS		Synthetic.		
XX				
FH	Key	Location/Qualifiers		
FT	Peptide	1..19		
FT	Protein	/note= "signal peptide"		
FT		20..2351		
FT	Region	/note= "mature Factor VIII:C"		
FT		20..1667		
FT	Modified-site	/note= "heavy chain fragment"		
FT		1667		
FT	Region	/label= Phe, Glu, Pro		
FT		1668..2350		
FT	Domain	/note= "light chain fragment"		
FT		760..1667		
XX		/note= "B domain"		
XX				
PN		N09J03195-A1.		
PD		30-JAN-1997.		
XX				
PE	09-JUL-1996;	96WO-US11444.		

[illegible]

QY	481	LIIFKNOASRPNIYPHGCTIDVRLYSRRLPKGYHLKDPRIIDGELFFKRYKPTVEDCP	540
Db	481	LIIFKNOASRPNIYPHGCTIDVRLYSRRLPKGYHLKDPRIIDGELFFKRYKPTVEDCP	540
QY	541	TKSDPRLCTRYSSSFVNMERDLASGLIGPLLCYKESYDORGNIMSDKRNWILFSVDE	600
Db	541	TKSDPRLCTRYSSSFVNMERDLASGLIGPLLCYKESYDORGNIMSDKRNWILFSVDE	600
QY	601	NRSWLTENIOFLFNPAGVOLDEPFOASNIMHSINGYFDSIQSLVCLHEHAYWILIS	660
Db	601	NRSWLTENIOFLFNPAGVOLDEPFOASNIMHSINGYFDSIQSLVCLHEHAYWILIS	660
QY	661	IGAOTDFLSVFPFGYTFKHKVYEDTLTLPFGSGTFVMSNENGLMIGCHNSPFRNG	720
Db	661	IGAOTDFLSVFPFGYTFKHKVYEDTLTLPFGSGTFVMSNENGLMIGCHNSPFRNG	720
QY	721	MTALLVSSCDKNIGDYEDSYEDISAYILSKNNAIEPRFSQNSRHPSTROKQFNATTI	780
Db	721	MTALLVSSCDKNIGDYEDSYEDISAYILSKNNAIEPRFSQNSRHPSTROKQFNATTI	780
QY	781	PENDIEKTDPMFAHRTMPKIONVSSDDLMLLRQSPTRHGLSLSDLOEAKYETESDDPS	840
Db	781	PENDIEKTDPMFAHRTMPKIONVSSDDLMLLRQSPTRHGLSLSDLOEAKYETESDDPS	840
QY	841	PGAIDSNNSLSEMTFRPOLHSGDMVFTPEGSLQLRLNEKLGTTAATELKILDFKVSST	900
Db	841	PGAIDSNNSLSEMTFRPOLHSGDMVFTPEGSLQLRLNEKLGTTAATELKILDFKVSST	900
QY	901	SNMLISTIPSDMLAAGDNTSILGPRSMVYHDSOLDTTLFGKSSPLTPSGGPISSIE	960
Db	901	SNMLISTIPSDMLAAGDNTSILGPRSMVYHDSOLDTTLFGKSSPLTPSGGPISSIE	960
QY	961	NDSKILBESGLANSOESWGKNVSTJESGRILFKGRAHGPALLTKDNALFKVYSILKTN	1020
Db	961	NDSKILBESGLANSOESWGKNVSTJESGRILFKGRAHGPALLTKDNALFKVYSILKTN	1020
QY	1021	KTSNNSATNRKTHIDGPSLTIENSPLYMONILIESDTEKKTPLIHDRLMOKNATATRL	1080
Db	1021	KTSNNSATNRKTHIDGPSLTIENSPLYMONILIESDTEKKTPLIHDRLMOKNATATRL	1080
QY	1081	NHMSNKTSSKNMEVVOQKRGPIPPDAONPMSFFKMLFLPESAHRIORTHGKNSLNSG	1140
Db	1081	NHMSNKTSSKNMEVVOQKRGPIPPDAONPMSFFKMLFLPESAHRIORTHGKNSLNSG	1140
QY	1141	QGPSKQOLVSLGPEKSVGQNFUSEKNKVVAGSGFTYDVLAKEMVPPSSRMLFTLNIDN	1200
Db	1141	QGPSKQOLVSLGPEKSVGQNFUSEKNKVVAGSGFTYDVLAKEMVPPSSRMLFTLNIDN	1200
QY	1201	LHENNTHNOEKKIOBELIEKKEKTLIOENVVLPQIHVTGTKNFMKNLFLSTRONVGSYD	1260
Db	1201	LHENNTHNOEKKIOBELIEKKEKTLIOENVVLPQIHVTGTKNFMKNLFLSTRONVGSYD	1260
QY	1261	GATAPVLODFRSLNNTNRTRKHAHFSKSGEENLBDLGNOTQOYIEKACTRISNPT	1320
Db	1261	GATAPVLODFRSLNNTNRTRKHAHFSKSGEENLBDLGNOTQOYIEKACTRISNPT	1320
QY	1321	SOONVTQSRKALKQFPLLEETELBKRIYDDTSTOWSKNMHLPSTLTIOIDYNEKE	1380
Db	1321	SOONVTQSRKALKQFPLLEETELBKRIYDDTSTOWSKNMHLPSTLTIOIDYNEKE	1380
QY	1381	KGAITOSPISDCLTSHSIRPOANRSPPLAKVSSPPIRPIYTLRVLFDONSSHLPAAZY	1440
Db	1381	KGAITOSPISDCLTSHSIRPOANRSPPLAKVSSPPIRPIYTLRVLFDONSSHLPAAZY	1440
QY	1441	RKDSGVQESSHFLQGAKNKMLSLATLEMGDORREVGSIGTATSVYKKEVENTVLP	1500
Db	1441	RKDSGVQESSHFLQGAKNKMLSLATLEMGDORREVGSIGTATSVYKKEVENTVLP	1500
QY	1501	KPDLPTSGKVELLKVYIYKODLFPETNSPCHLDLVSGLIOTCBATIKMNAARP	1560
Db	1501	KPDLPTSGKVELLKVYIYKODLFPETNSPCHLDLVSGLIOTCBATIKMNAARP	1560

QY	1561	GVPEFLVATESAKPSPSKILDPILAMDNHSTOIPREEMKSOEKSPEKATKKKDTILS	1620
Db	1561	GVPEFLVATESAKPSPSKILDPILAMDNHSTOIPREEMKSOEKSPEKATKKKDTILS	1620
QY	1621	MACESNHAIAAINEGONKPEIETVNAKQRTRELCSQNPVLRKHORETTRTLOSDOE	1680
Db	1621	MACESNHAIAAINEGONKPEIETVNAKQRTRELCSQNPVLRKHORETTRTLOSDOE	1680
QY	1681	IDYDDTISEVKKEDFDIYDEDENOSSPSPFOKTRHVFIAAVERLMDYGMSSPHVLRN	1740
Db	1681	IDYDDTISEVKKEDFDIYDEDENOSSPSPFOKTRHVFIAAVERLMDYGMSSPHVLRN	1740
QY	1741	AOSGSVPQFKVYQOEFTOSFPQIYVGEINELGLIPYIAAEVENIMVTRRNASR	1800
Db	1741	AOSGSVPQFKVYQOEFTOSFPQIYVGEINELGLIPYIAAEVENIMVTRRNASR	1800
QY	1801	PISTFSSLSIYEEOROGAPRKNTYKPRETKTYFMKVQHMAPTDEFCANAFSDV	1860
Db	1801	PISTFSSLSIYEEOROGAPRKNTYKPRETKTYFMKVQHMAPTDEFCANAFSDV	1860
QY	1861	DLEKDVHSGILIGPLVCHTWTLPAGHQVTVQEFALFTIFDETSKWTETENNERCRA	1920
Db	1861	DLEKDVHSGILIGPLVCHTWTLPAGHQVTVQEFALFTIFDETSKWTETENNERCRA	1920
QY	1921	PCNIOMEDPTEKENYRPHAINGYIMDTLPGLVMAODRIMVYLLSNGSNENHSHIFSGH	1980
Db	1921	PCNIOMEDPTEKENYRPHAINGYIMDTLPGLVMAODRIMVYLLSNGSNENHSHIFSGH	1980
QY	1981	VFTVYKKEEYKMAIYNLYPGVETVEMLPKSAAGIYRVECLIGHLHAGNSTLFLYTSNKC	2040
Db	1981	VFTVYKKEEYKMAIYNLYPGVETVEMLPKSAAGIYRVECLIGHLHAGNSTLFLYTSNKC	2040
QY	2041	QTPPLMASGHIRDFQITASGOYGOAMPKILARLHSGSINAMSTKEPFSWIKVDLAPMII	2100
Db	2041	QTPPLMASGHIRDFQITASGOYGOAMPKILARLHSGSINAMSTKEPFSWIKVDLAPMII	2100
QY	2101	HGIKTOGARQFSSLYISQFTIYSLDGKKNQOTRGSTGLTAVFFGANDVSSGIKHNIFN	2160
Db	2101	HGIKTOGARQFSSLYISQFTIYSLDGKKNQOTRGSTGLTAVFFGANDVSSGIKHNIFN	2160
QY	2161	PLIARIYRLRPHYSTRSLRBEMLGCDLNSGAMPKAGMSKAIISNOQTASVFTNFA	2220
Db	2161	PLIARIYRLRPHYSTRSLRBEMLGCDLNSGAMPKAGMSKAIISNOQTASVFTNFA	2220
QY	2221	TWSPSKARLHLQGRSNAMRPQVNNPKEMLQVDFOKTKRVGVTTOGVKSILTSMYKREFL	2280
Db	2221	TWSPSKARLHLQGRSNAMRPQVNNPKEMLQVDFOKTKRVGVTTOGVKSILTSMYKREFL	2280
QY	2281	ISSSDQHQWTLFQONGKVVVEFGNDSFTPVVNSLDPPLTFRYLRIHFQSVVHQLALRM	2340
Db	2281	ISSSDQHQWTLFQONGKVVVEFGNDSFTPVVNSLDPPLTFRYLRIHFQSVVHQLALRM	2340
QY	2341	EVLGCEADQDLY 2351	
Db	2341	EVLGCEADQDLY 2351	

RESULT 11
AAM11425
ID AAM11425 standard; Protein: 2351 aa.
XX
XX AAM11425;
XX
XX 20-NOV-1997 (first entry)
XX
XX Active Factor VIII:C analogue T1308X.
XX
XX Factor VIII:C: analogue; glycoprotein; blood coagulation cascade;
XX fibrinogen; fibrin clot; haemostasis; haemophilia A; bleeding diathesis;
XX plasma protease; thrombin; immunogen; antibody; haemophiliaac; therapy;
XX proteolytic cleavage.
XX
XX Homo sapiens.
OS

OS Synthetic.
XX Key Location/Qualifiers
FH Peptide 1..19 /note="signal peptide"
FT Protein 20..2351 /note="mature Factor VIII:C"
FT Region 20..1667 /note="heavy chain fragment"
FT Modified-site 1327 /label="Phe, Glu, Pro
FT Region 1668..2350 /note="light chain fragment"
FT Domain 760..1667 /note="B domain"
XX
XX
XX NO9703195-A1.
XX
XX 30-JAN-1997.
XX
XX 09-JUL-1996; 96WO-US11444.
XX
XX 11-JUL-1995; 95US-0001025.
XX
XX (CHIR) CHIRON CORP.
XX
XX Cohen FE, Hung DF, Innis M;
XX WPI; 1997-119050/11.
XX
XX Factor VIII:C analog modified adjacent to a non-activating Arg
PT residue - used in the treatment of haemophilias, by improvement of
PT haemostasis
XX
XX
XX Claim 28; Page -; 90pp; English.
XX
XX AAM1330-W11472 represent active Factor VIII:C analogues of the
CC invention. These sequences were created by mutating the wild type Factor
CC VIII:C coding sequence (see AAT31357) using mutagenic primers. The
CC analogues comprise a native Factor VIII:C polypeptide modified at a site
CC adjacent to a non-activating Arg residue so that a Arg-Pro or Pro-Arg
CC dipeptide is created. Factor VIII:C is a large glycoprotein that
CC participates in the blood coagulation cascade that ultimately converts
CC soluble fibrinogen to insoluble fibrin clot, effecting haemostasis. A
CC deficiency in Factor VIII:C is responsible for haemophilia A, which is an
CC X-chromosome-linked inherited bleeding diathesis. Factor VIII:C is
CC activated by plasma proteases, such as thrombin. During activation the
CC mature polypeptide is cleaved to generate heavy and light chain fragments
CC that are further cleaved. Complexes of two or more of the analogues,
CC nucleic acids and vectors encoding them may be used alone or in
CC conjunction with each other, for the prevention or treatment of active
CC Factor VIII:C deficiency in a mammal. The analogues may be used as
CC immunogens to raise antibodies, and in the treatment of hemophilias, by
CC improvement of haemostasis. The analogues are resistant to proteolytic
CC cleavage and display increased plasma half-life. They may be administered
CC at lower dosages and by different modes of administration.
XX
XX
SQ Sequence 2351 AA;
Query Match 100.0%; Score 12413; DB 18; Length 2351;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2350; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 MOETLSTCFPLCLIRFCFSATRRYYIGAVELSMOYSDGELPVAPRPVPKSPFFN 60
DB 1 MOETLSTCFPLCLIRFCFSATRRYYIGAVELSMOYSDGELPVAPRPVPKSPFFN 60
QY 61 TSYVYKTKLTFVEFTDHLFNIAKRPMPKGLGPTIOAEVTDYVITLKMAASHVSLAV 120
DB 61 TSYVYKTKLTFVEFTDHLFNIAKRPMPKGLGPTIOAEVTDYVITLKMAASHVSLAV 120
QY 121 GVSYNKASGAEYDQTSOREKEDKVFPGSGSHTYVMQVLKENGPMASDPLCLTYSYLSH 180
DB 121 GVSYNKASGAEYDQTSOREKEDKVFPGSGSHTYVMQVLKENGPMASDPLCLTYSYLSH 180

DB 121 GVSYNKASGAEYDQTSOREKEDKVFPGSGSHTYVMQVLKENGPMASDPLCLTYSYLSH 180
QY 161 VDLVKDNLNSLIGALLVCREGSLAKETQTLHFKFLTFVPEDEKSMHSETKNSLQDDB 240
DB 161 VDLVKDNLNSLIGALLVCREGSLAKETQTLHFKFLTFVPEDEKSMHSETKNSLQDDB 240
QY 241 AASARAPKPKHTYVNGVNRSLPGLIGHRKSYVWYIGMGTPEVHSIFEGHTPLVNRH 300
DB 241 AASARAPKPKHTYVNGVNRSLPGLIGHRKSYVWYIGMGTPEVHSIFEGHTPLVNRH 300
QY 301 KQASLEISPTFLTAOTLLMDLGOFLFCHISSHQHGMEAYVAYKVSCEPRLRMKNE 360
DB 301 KQASLEISPTFLTAOTLLMDLGOFLFCHISSHQHGMEAYVAYKVSCEPRLRMKNE 360
QY 361 EAEDYDDDLTDSFMDVVRPDDNSPSFIQIRSAKHKPKTWVHYIAAEEDMDYAPLVLA 420
DB 361 EAEDYDDDLTDSFMDVVRPDDNSPSFIQIRSAKHKPKTWVHYIAAEEDMDYAPLVLA 420
QY 421 PDDRSYSQYLNNGPQRIQKTKYKPMATDETETKRAIGHESGLIGPLVGEVDTL 480
DB 421 PDDRSYSQYLNNGPQRIQKTKYKPMATDETETKRAIGHESGLIGPLVGEVDTL 480
QY 481 LIIFRNOASRPYNIYPHGITTDRPLYSRRLPKGKHLKDFPLPGLFIKKWTYVEDP 540
DB 481 LIIFRNOASRPYNIYPHGITTDRPLYSRRLPKGKHLKDFPLPGLFIKKWTYVEDP 540
QY 541 TKSPDCLTRYSSPFNMERDLASGLIGPLIICYSKSVORONQIMSDKRVILFVDE 600
DB 541 TKSPDCLTRYSSPFNMERDLASGLIGPLIICYSKSVORONQIMSDKRVILFVDE 600
QY 601 NRSWTLFENIQRLFPAPAGVQLEDEPQASVHNSINGVPEPSLOLSLCHVAYVYIS 660
DB 601 NRSWTLFENIQRLFPAPAGVQLEDEPQASVHNSINGVPEPSLOLSLCHVAYVYIS 660
QY 661 IGAQDFLSVFFSGYTFKKRWYEDTLTFPEFSGETVYMSMNPGLMILGCHNSDERNG 720
DB 661 IGAQDFLSVFFSGYTFKKRWYEDTLTFPEFSGETVYMSMNPGLMILGCHNSDERNG 720
QY 721 MTALTKYSCDKNTGYEDSYEDISATLKSNNALTEPSSQNSRHSPTROKOFNATY 780
DB 721 MTALTKYSCDKNTGYEDSYEDISATLKSNNALTEPSSQNSRHSPTROKOFNATY 780
QY 781 PENDEKTDMPFAHRPMPKIONVSSDLMLRQSPTRHGSLSDLOEAKYEFSPDD 840
DB 781 PENDEKTDMPFAHRPMPKIONVSSDLMLRQSPTRHGSLSDLOEAKYEFSPDD 840
QY 841 PGALDSNNSLSSEWTFRRPOLBHSQDMFTPESSGLQRLNKEIGTAATELKKLDKYSST 900
DB 841 PGALDSNNSLSSEWTFRRPOLBHSQDMFTPESSGLQRLNKEIGTAATELKKLDKYSST 900
QY 901 SNMLISTIPSDNLAAGTDNTSLAGPSPMPVHYDSQDLDTLFGKSSPPTESGPLSLSEE 960
DB 901 SNMLISTIPSDNLAAGTDNTSLAGPSPMPVHYDSQDLDTLFGKSSPPTESGPLSLSEE 960
QY 961 NNDKLLSGLMSQSSSGKKNVSTESGRLFKKRAHAGPALLTDNMLFVSTSLKTN 1020
DB 961 NNDKLLSGLMSQSSSGKKNVSTESGRLFKKRAHAGPALLTDNMLFVSTSLKTN 1020
QY 1021 KTSNNSATNRKTHIDGSLIENSQVQNLISDETFEKKVPTLJHDMKADKATPLRL 1080
DB 1021 KTSNNSATNRKTHIDGSLIENSQVQNLISDETFEKKVPTLJHDMKADKATPLRL 1080
QY 1081 NMSNKTSSKNMEVQOKKEGPIPPDAQNDMSFFKMLFLPESARWIOPTHGKNSLNSG 1140
DB 1081 NMSNKTSSKNMEVQOKKEGPIPPDAQNDMSFFKMLFLPESARWIOPTHGKNSLNSG 1140
QY 1141 QGSPKQVLSLGPESVSEONPLSEKNVYVGKEFLKVDGLKEMVPSRNPLTLMDN 1200
DB 1141 QGSPKQVLSLGPESVSEONPLSEKNVYVGKEFLKVDGLKEMVPSRNPLTLMDN 1200
QY 1201 LHENNTNQEKKIQEIEKEKTELIOENVVLPQIHVTYGTKNFKMLFLSLTRONVBSYD 1260
DB 1201 LHENNTNQEKKIQEIEKEKTELIOENVVLPQIHVTYGTKNFKMLFLSLTRONVBSYD 1260

```

OY 1261 GAYAPVLDPFRSLNDSTNRTKHTAHFSKKEEENLGLGNQTKOJVEKYACTTRISPT 1320
DB 1261 GAYAPVLDPFRSLNDSTNRTKHTAHFSKKEEENLGLGNQTKOJVEKYACTTRISPT 1320
OY 1321 SOONFVTOXSKRALKOPRLPLEETELERKIYVDPTSTOMSKNNKHLTPSTLQIDYENKE 1380
DB 1321 SOONFVTOXSKRALKOPRLPLEETELERKIYVDPTSTOMSKNNKHLTPSTLQIDYENKE 1380
OY 1381 KGATVOSPLSDCLTRSHSTIPQANRSPPLAKYSSPSPSTIRPLTRVLFDNSHLPAASY 1440
DB 1381 KGATVOSPLSDCLTRSHSTIPQANRSPPLAKYSSPSPSTIRPLTRVLFDNSHLPAASY 1440
OY 1441 RKDSGVQESSHFLQAKKNNLSLAILTLEMTDQREVGSIGTSAATSVYTKRYENVLP 1500
DB 1441 RKDSGVQESSHFLQAKKNNLSLAILTLEMTDQREVGSIGTSAATSVYTKRYENVLP 1500
OY 1501 KPDLPTSGKVELLPKVHIYQKDLFPTETSNQSPGHLDLVEGSLQGTGGAIKMNEANRP 1560
DB 1501 KPDLPTSGKVELLPKVHIYQKDLFPTETSNQSPGHLDLVEGSLQGTGGAIKMNEANRP 1560
OY 1561 GKVPFLRVATESSAKTSPSKLLDPLANDNHYGTQIPKEEMKSQSEKPEKFAFKKDTITSL 1620
DB 1561 GKVPFLRVATESSAKTSPSKLLDPLANDNHYGTQIPKEEMKSQSEKPEKFAFKKDTITSL 1620
OY 1621 NACESNHATAINEONKPELEVMAKQRTBRLCSQNPVLRKHORELTTRTLOSOOE 1680
DB 1621 NACESNHATAINEONKPELEVMAKQRTBRLCSQNPVLRKHORELTTRTLOSOOE 1680
OY 1681 IDYDPTISVEAKKEDFDIYDENOSPSRFOKTRHFFIAVERLMDYGMSSPHYVLRNR 1740
DB 1681 IDYDPTISVEAKKEDFDIYDENOSPSRFOKTRHFFIAVERLMDYGMSSPHYVLRNR 1740
OY 1741 AQSGSVPOFKKVVFOEFTDSFTQPLYRGELNEHLGLGPYRAEVEDNIMVTFRNOASR 1800
DB 1741 AQSGSVPOFKKVVFOEFTDSFTQPLYRGELNEHLGLGPYRAEVEDNIMVTFRNOASR 1800
OY 1801 PYSFYSLSLYEEDROGAEPKRNFKYKPNETKTYFMKVOHMAPTKDEDFCANKAYFSDV 1860
DB 1801 PYSFYSLSLYEEDROGAEPKRNFKYKPNETKTYFMKVOHMAPTKDEDFCANKAYFSDV 1860
OY 1861 DLEKDVHSLIGPLLYCHTNTLNPARGROYVQOEALFTTIEDTSTSMPTENMERNRA 1920
DB 1861 DLEKDVHSLIGPLLYCHTNTLNPARGROYVQOEALFTTIEDTSTSMPTENMERNRA 1920
OY 1921 PCNIQMEDPTFKENTRFALINGYIMDTLPGLVMAQDORIRMYLLSMGSMENHSHSPSGH 1980
DB 1921 PCNIQMEDPTFKENTRFALINGYIMDTLPGLVMAQDORIRMYLLSMGSMENHSHSPSGH 1980
OY 1981 VFTVRRKKEEYKMAALNYLPGVFETVEMLPKAGIWRVCLIGELHAGNSTLEFLVSNKC 2040
DB 1981 VFTVRRKKEEYKMAALNYLPGVFETVEMLPKAGIWRVCLIGELHAGNSTLEFLVSNKC 2040
OY 2041 QTPLGMAAGHINDPQITASGOYGOMAPKLARLHYGSGINASTKEPSPMKDLPAMI 2100
DB 2041 QTPLGMAAGHINDPQITASGOYGOMAPKLARLHYGSGINASTKEPSPMKDLPAMI 2100
OY 2101 HGKITGAKOKSSLYISQFTIMYSIDGKKNQTYGNSGTGLMFRGNDSSGISHNIFN 2160
DB 2101 HGKITGAKOKSSLYISQFTIMYSIDGKKNQTYGNSGTGLMFRGNDSSGISHNIFN 2160
OY 2161 PPIIARTIRLHPTRHSINSTLRLMELMGCDLNSCSPMLCMEKASISDAQITASSYTFNFA 2220
DB 2161 PPIIARTIRLHPTRHSINSTLRLMELMGCDLNSCSPMLCMEKASISDAQITASSYTFNFA 2220
OY 2221 TWSPSKARLHLQGRSNAMRPOVNNPKEMLVDFOFKTMKVTYVTOGVKSLTSMYKFEFL 2280
DB 2221 TWSPSKARLHLQGRSNAMRPOVNNPKEMLVDFOFKTMKVTYVTOGVKSLTSMYKFEFL 2280
OY 2281 ISSSQDHOHWTLFPQNGKVVYVQGNODSPFPVNVSLDPLLTRLRLRHPQSWHQAIDRM 2340
DB 2281 ISSSQDHOHWTLFPQNGKVVYVQGNODSPFPVNVSLDPLLTRLRLRHPQSWHQAIDRM 2340

```

```

OY 2341 EYLCEAODLY 2351
DB 2341 EYLCEAODLY 2351

RESULT 12
AA011419
ID AA011419 standard; protein; 2351 AA.
XX
AC AA011419;
XX
DF 20-NOV-1997 (first entry)
XX
DE Active Factor VIII:C analogue, delta 1311, + Pro Insertion.
XX
KW Factor VIII:C; analogue; glycoprotein; blood coagulation cascade;
KW fibrinogen; fibrin clot; haemostasis; haemophilia A; bleeding diathesis;
KW plasma protease; thrombin; immunogen; antibody; haemophilic; therapy;
KW proteolytic cleavage.
XX
OS Homo sapiens.
OS Synthetic.
XX
FH Key 1..19 Location/Qualifiers
FH Peptide /note= "signal peptide"
FH Protein /note= "2351"
FH Region /note= "mature Factor VIII:C"
FH Region /note= "1667"
FH Misc-difference /note= "heavy chain fragment"
FH FT 1329..1330
FH FT /note= "site of 1 residue deletion"
FH FT Modified-site 1330
FH FT /note= "inserted residue"
FH FT Region 1668..2350
FH FT /note= "light chain fragment"
FH FT Domain 760..1667
FH FT /note= "B domain"
XX
XX MO9703195-AL.
XX
XX 30-JAN-1997.
XX
XX 09-JUL-1996; 96MO-0511444.
XX
XX 11-JUL-1995; 95US-0001025.
XX
XX (CHIR ) CHIRON CORP.
XX
XX Cohen FE, Hung DT, Innis M;
XX
XX WPI: 1997-119050/11.
XX
XX Factor VIII:C analog modified adjacent to a non-activating Arg
XX residue - used in the treatment of haemophilias, by improvement of
XX hemostasis
XX
XX Claim 27; Page -: 90pp; English.
XX
XX AA011330-W11472 represent active Factor VIII:C analogues of the
XX invention. These sequences were created by mutating the wild type Factor
XX VIII:C coding sequence (see AA051357) using mutagenic primers. The
XX analogues comprise a native Factor VIII:C polypeptide modified at a site
XX adjacent to a non-activating Arg residue so that a Arg-Pro or Pro-Arg
XX dipeptide is created. Factor VIII:C is a large glycoprotein that
XX participates in the blood coagulation cascade that ultimately converts
XX soluble fibrinogen to insoluble fibrin clot, effecting hemostasis. A
XX deficiency in Factor VIII:C is responsible for haemophilia A, which is an
XX X-chromosome-linked inherited bleeding diathesis. Factor VIII:C is
XX activated by plasma proteases, such as thrombin. During activation the
XX mature polypeptide is cleaved to generate heavy and light chain fragments
XX that are further cleaved. Complexes of two or more of the analogues,
XX nucleic acids and vectors encoding them may be used alone or in

```


CC conjunction with each other, for the prevention or treatment of active
 CC Factor VIII:C deficiency in a mammal. The analogues may be used as
 CC immunogens to raise antibodies, and in the treatment of haemophilias, by
 CC improvement of haemostasis. The analogues are resistant to proteolytic
 CC cleavage and display increased plasma half-life. They may be administered
 CC at lower dosages and by different modes of administration.

xx Sequence 2351 AA:

Query Match 100.0%; Score 12413; DB 18; Length 2351;

Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2350; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MOELSTCFELCLRCFSATRRYTLGAVELSWPTMQSDGELPVDARPPRVKSPFN 60
 Db 1 MOELSTCFELCLRCFSATRRYTLGAVELSWPTMQSDGELPVDARPPRVKSPFN 60
 Qy 61 TSVYTKTLFEFDLHFNIAKPRPMGGLGPTIOAEYTDVYITLKNASHVSLHAY 120
 Db 61 TSVYTKTLFEFDLHFNIAKPRPMGGLGPTIOAEYTDVYITLKNASHVSLHAY 120
 Qy 121 GVSYKASGAEYDDOTSQREKEDKVPFGSSHTYVQVLKENGPMASDPLCLTYSLSH 180
 Db 121 GVSYKASGAEYDDOTSQREKEDKVPFGSSHTYVQVLKENGPMASDPLCLTYSLSH 180
 Qy 181 VDLVNDLSGLIGALVCRGSLAKEKTOITLHKFTLLFAVFDGSKMSHSEKNSIMQDRD 240
 Db 181 VDLVNDLSGLIGALVCRGSLAKEKTOITLHKFTLLFAVFDGSKMSHSEKNSIMQDRD 240
 Qy 241 AASARAPKMTWNGVYNSLPGIGCRKSVYHVGMTPEVHSFTLSGHEFLVRNH 300
 Db 241 AASARAPKMTWNGVYNSLPGIGCRKSVYHVGMTPEVHSFTLSGHEFLVRNH 300
 Qy 301 ROASLEISPTFTLTAOTLLMDJGFLFCGHSQHODGKEAYKVDSCPEEPQALMKANE 360
 Db 301 ROASLEISPTFTLTAOTLLMDJGFLFCGHSQHODGKEAYKVDSCPEEPQALMKANE 360
 Qy 361 EAEYDDDLTDEMDVVRFPDDNSPSFIOIRSVAKKHPTWHTIAEEDMDYAPLVLA 420
 Db 361 EAEYDDDLTDEMDVVRFPDDNSPSFIOIRSVAKKHPTWHTIAEEDMDYAPLVLA 420
 Qy 421 PDPRSASQOYLNGFPORIGRKYKVFMAAYTDEFKTRREALIQHESGILGPLLGYVGDTL 480
 Db 421 PDPRSASQOYLNGFPORIGRKYKVFMAAYTDEFKTRREALIQHESGILGPLLGYVGDTL 480
 Qy 481 LIIFKNOASRPYNTYPRGIDVAPLYSRRLPKGYKHLKDPILLGPIPKYKWTYVVDGP 540
 Db 481 LIIFKNOASRPYNTYPRGIDVAPLYSRRLPKGYKHLKDPILLGPIPKYKWTYVVDGP 540
 Qy 541 TKSDFRCLTRYSSFVNMERDLASGLIGPLICTKESVDORGNOIMSKRNYLFSVDE 600
 Db 541 TKSDFRCLTRYSSFVNMERDLASGLIGPLICTKESVDORGNOIMSKRNYLFSVDE 600
 Qy 601 NRSWYLTENIORFLPNAGVQLEDEPEQASNTMISNGVYFDSIQSLVCLHEVAYWYILS 660
 Db 601 NRSWYLTENIORFLPNAGVQLEDEPEQASNTMISNGVYFDSIQSLVCLHEVAYWYILS 660
 Qy 661 IGAQTDPLSVFSGYTFKHKMYEDTLTLPFSGEYVPMENRPMGMLIGCNSDFNRNG 720
 Db 661 IGAQTDPLSVFSGYTFKHKMYEDTLTLPFSGEYVPMENRPMGMLIGCNSDFNRNG 720
 Qy 721 MTALLKVSSCDKNTGDIYEDSYDLSAYILSKNNAIEPNSFSONSRHSPTRKOPNATTI 780
 Db 721 MTALLKVSSCDKNTGDIYEDSYDLSAYILSKNNAIEPNSFSONSRHSPTRKOPNATTI 780
 Qy 781 PENDIEKTDPMFAHRTMPRIQWVSSDMLMLROSPTPHGLSLSDQAKETFSDDPS 840
 Db 781 PENDIEKTDPMFAHRTMPRIQWVSSDMLMLROSPTPHGLSLSDQAKETFSDDPS 840
 Qy 841 PGALISNNISLSEMTHTFHPOLAHSGDNVTFPESGQIOLRNLKIGTATATLKKLDKVSST 900
 Db 841 PGALISNNISLSEMTHTFHPOLAHSGDNVTFPESGQIOLRNLKIGTATATLKKLDKVSST 900

Qy 901 SNNLITIPSDMLAGTNDTSSLGPPSPMPVHDSOLDTTLRGKSSPIRESGGPLTSEE 960
 Db 901 SNNLITIPSDMLAGTNDTSSLGPPSPMPVHDSOLDTTLRGKSSPIRESGGPLTSEE 960
 Qy 961 NNDKSLIESGLMNSOESSWGKNVSTESGRLKRGKRAKAPLLTKDNALFKYISILKTN 1020
 Db 961 NNDKSLIESGLMNSOESSWGKNVSTESGRLKRGKRAKAPLLTKDNALFKYISILKTN 1020
 Qy 1021 KTSNNSATNRKTHIDGPSLLIENSVMONILSDETERKVTPLIHDRMLDMKNATLRL 1080
 Db 1021 KTSNNSATNRKTHIDGPSLLIENSVMONILSDETERKVTPLIHDRMLDMKNATLRL 1080
 Qy 1081 NMSNKTSTSKNMEVYQKKEGPIPPDAQPPDMSPFKILFIPESARWIOPTHGKNSLNG 1140
 Db 1081 NMSNKTSTSKNMEVYQKKEGPIPPDAQPPDMSPFKILFIPESARWIOPTHGKNSLNG 1140
 Qy 1141 QGSPKQVSLGPEKSVGQNFLEKKNVYVGKGFPTDVGLEKMPSPSRNLFILNIDN 1200
 Db 1141 QGSPKQVSLGPEKSVGQNFLEKKNVYVGKGFPTDVGLEKMPSPSRNLFILNIDN 1200
 Qy 1201 LHENNTNDEKTIQEBIEKKTLLIQENVVLPQIHVYTGKKNPKNLFILSTRONVBSGD 1260
 Db 1201 LHENNTNDEKTIQEBIEKKTLLIQENVVLPQIHVYTGKKNPKNLFILSTRONVBSGD 1260
 Qy 1261 GAYAPVLODFRSINDSTRKTHAHSKKEEENLEGLAQOTQIYEKACTTRISPT 1320
 Db 1261 GAYAPVLODFRSINDSTRKTHAHSKKEEENLEGLAQOTQIYEKACTTRISPT 1320
 Qy 1321 SQONFVTOBSKRALQOFRLPELELEKRTIYDDTSQWKNMHLPPSLTIOIDVNEKE 1380
 Db 1321 SQONFVTOBSKRALQOFRLPELELEKRTIYDDTSQWKNMHLPPSLTIOIDVNEKE 1380
 Qy 1381 KGATVQSPDLCLTNSHSTPOANSRSPPLAKVSPSPSTIRYTLRVLFONSHPLPAASY 1440
 Db 1381 KGATVQSPDLCLTNSHSTPOANSRSPPLAKVSPSPSTIRYTLRVLFONSHPLPAASY 1440
 Qy 1441 RKDQSVQESSHFLQAKKNLSLAILTLEMTDQDQREVSGLTSATNSVTYKKVENVLP 1500
 Db 1441 RKDQSVQESSHFLQAKKNLSLAILTLEMTDQDQREVSGLTSATNSVTYKKVENVLP 1500
 Qy 1501 KPDLPTKSGVELLPVHVIYOKDLPETPENSQPHLIDLVEGSLAQTEGAIKMNEANRP 1560
 Db 1501 KPDLPTKSGVELLPVHVIYOKDLPETPENSQPHLIDLVEGSLAQTEGAIKMNEANRP 1560
 Qy 1561 GKVPFLVATESSAKTPSKLDPILAMNHNHGTQIPREEMKSOEKSPKTAFFKKKDTILSL 1620
 Db 1561 GKVPFLVATESSAKTPSKLDPILAMNHNHGTQIPREEMKSOEKSPKTAFFKKKDTILSL 1620
 Qy 1621 NACSSNNAIAITNEGONRPEIETVMAOGSTERLCSQNPVYLAKHOREITRTLOSODE 1680
 Db 1621 NACSSNNAIAITNEGONRPEIETVMAOGSTERLCSQNPVYLAKHOREITRTLOSODE 1680
 Qy 1681 IDYDQTSVEMKKEDEYIDEDENOSPRSFOKTRHYETIAAVERLMDYGNSSPHVLRNR 1740
 Db 1681 IDYDQTSVEMKKEDEYIDEDENOSPRSFOKTRHYETIAAVERLMDYGNSSPHVLRNR 1740
 Qy 1741 AOGSGVPOFKKVVQFETGSPQPLVREGELNHLGLCPYRAEVEDNIMVYFRNOASR 1800
 Db 1741 AOGSGVPOFKKVVQFETGSPQPLVREGELNHLGLCPYRAEVEDNIMVYFRNOASR 1800
 Qy 1801 PYSTYSSLIEYEDQOQAGPRKKNVPRNETKTYFKFKVQHNNAPRKQEPCKMAAFSDV 1860
 Db 1801 PYSTYSSLIEYEDQOQAGPRKKNVPRNETKTYFKFKVQHNNAPRKQEPCKMAAFSDV 1860
 Qy 1861 DLEKDVHSGLIGPLVCHTNTLNPAGHROYTVOEFALPFTJFDETKSMYFTEMMERNCA 1920
 Db 1861 DLEKDVHSGLIGPLVCHTNTLNPAGHROYTVOEFALPFTJFDETKSMYFTEMMERNCA 1920
 Qy 1921 PCNTOMDPTFEKENYFFAHINGYIMDTLPGLVAAQOPRIRWLLLSMGSNENISIFPSGH 1980
 Db 1921 PCNTOMDPTFEKENYFFAHINGYIMDTLPGLVAAQOPRIRWLLLSMGSNENISIFPSGH 1980
 Qy 1981 VFTYRKKEEYKMAVLYLFGVYETVEMLPKKAQINRVECLIOEHLHAGMSTLPVLYSNKC 2040

Db	541	TKSDPRCLTRYSSVFMEDLASGLIGPLLCYKESVDQRGNQISDKRNVILFESVDE	1680
Qy	601	NRSWLTENIORPLPNAGVLEDPFQASNMISNGVYFDSLOISVCLHVAWYILS	1740
Db	601	NRSWLTENIORPLPNAGVLEDPFQASNMISNGVYFDSLOISVCLHVAWYILS	1740
Qy	661	IGAOTDFLSVFSGYFHKMYEDTLFPFSGEYFNSMENPGWIIIGCHNSDFNRNG	1800
Db	661	IGAOTDFLSVFSGYFHKMYEDTLFPFSGEYFNSMENPGWIIIGCHNSDFNRNG	1800
Qy	721	MTALLKSSCDKNAGDYEDSYEDISAYLLSKNNAIEPRSFQSNRHPSTROKOFNATTI	1860
Db	721	MTALLKSSCDKNAGDYEDSYEDISAYLLSKNNAIEPRSFQSNRHPSTROKOFNATTI	1860
Qy	781	PENDLEKTDWFAHRTPMPKIONVSSDLMLNQSPTPHGLSLSLOAKETFSDDPS	1920
Db	781	PENDLEKTDWFAHRTPMPKIONVSSDLMLNQSPTPHGLSLSLOAKETFSDDPS	1920
Qy	841	PGALDSNNSLSEKMTFRPQLHSGDMVTTPESGQLRLNEKLTATATLKLKLPKYST	1980
Db	841	PGALDSNNSLSEKMTFRPQLHSGDMVTTPESGQLRLNEKLTATATLKLKLPKYST	1980
Qy	901	SNMLISTIPSDNLAAGDNTSSLGPPSMFVHDSQDITTFGKSSSPLTESGSPLSLEE	2040
Db	901	SNMLISTIPSDNLAAGDNTSSLGPPSMFVHDSQDITTFGKSSSPLTESGSPLSLEE	2040
Qy	961	NDSKILLESGLMNSQESSMGKNVSTESGRLFKGRAGHAPALLTKONALFKYSILKTN	2100
Db	961	NDSKILLESGLMNSQESSMGKNVSTESGRLFKGRAGHAPALLTKONALFKYSILKTN	2100
Qy	1021	KTSNNSATNRTKHIDGSLLENSPSWONILLESSTPEKKVYPLIHDRMLMKNTAARL	2160
Db	1021	KTSNNSATNRTKHIDGSLLENSPSWONILLESSTPEKKVYPLIHDRMLMKNTAARL	2160
Qy	1081	NHMSNKTSSKRMEMVQOKKEGPIPPDAQNPDMSEFKMLFPESAHWIORTHGKNSLNG	2220
Db	1081	NHMSNKTSSKRMEMVQOKKEGPIPPDAQNPDMSEFKMLFPESAHWIORTHGKNSLNG	2220
Qy	1141	QGSPPKOLVSLGPEKSEVGNFLSEKNKVVYVKGFTKDVKLKMYPPSSRLNLTINDN	2280
Db	1141	QGSPPKOLVSLGPEKSEVGNFLSEKNKVVYVKGFTKDVKLKMYPPSSRLNLTINDN	2280
Qy	1201	LHENTNHOEKRIQEBIEKEKTELIOENVLPQIHVYVGTCKMKMNFILSTQONEGSYD	2340
Db	1201	LHENTNHOEKRIQEBIEKEKTELIOENVLPQIHVYVGTCKMKMNFILSTQONEGSYD	2340
Qy	1261	GAYAPVLODFRSLNDSTNRTKHTAHFSKKGEEENLEGLNQTKQIYERYACTRISBNT	2400
Db	1261	GAYAPVLODFRSLNDSTNRTKHTAHFSKKGEEENLEGLNQTKQIYERYACTRISBNT	2400
Qy	1321	SOONFYTORSKRALKOPFLPLEETELEKRIYDDTSTQMSKNMKHLPPSTLQIDVNEKE	2460
Db	1321	SOONFYTORSKRALKOPFLPLEETELEKRIYDDTSTQMSKNMKHLPPSTLQIDVNEKE	2460
Qy	1381	KGAITQSPLSOCTLRSHSIPQANRSPPLAKVSSSPSTIRPYLRLVLFQONSHLPAASY	2520
Db	1381	KGAITQSPLSOCTLRSHSIPQANRSPPLAKVSSSPSTIRPYLRLVLFQONSHLPAASY	2520
Qy	1441	RKDSGVQSSHFLQAGAKNNLSLAILLEMTGDDRREVSLGTSATNSVYTKKVENTVLP	2580
Db	1441	RKDSGVQSSHFLQAGAKNNLSLAILLEMTGDDRREVSLGTSATNSVYTKKVENTVLP	2580
Qy	1501	KPDLPKTSQKVELLKKVITQKOLFPETNSNGSPGHLDVYGSLLQOTBEAIIWVNAANP	2640
Db	1501	KPDLPKTSQKVELLKKVITQKOLFPETNSNGSPGHLDVYGSLLQOTBEAIIWVNAANP	2640
Qy	1561	GKVPFLRAVATESAATPSKLLDPLAMDNHGTQIQPREKMSQEKSPERKAFKKKDTIISL	2700
Db	1561	GKVPFLRAVATESAATPSKLLDPLAMDNHGTQIQPREKMSQEKSPERKAFKKKDTIISL	2700
Qy	1621	NACESNHAIAINEGONKPEIEVWAKOGRTERLCSQNPVILKRHOIREITRTILOSDOE	2760
Db	1621	NACESNHAIAINEGONKPEIEVWAKOGRTERLCSQNPVILKRHOIREITRTILOSDOE	2760
Db	1621	NACESNHAIAINEGONKPEIEVWAKOGRTERLCSQNPVILKRHOIREITRTILOSDOE	1680
Qy	1681	IDYDPTLSVEKKEDDPDIYEDENQSPRSFQKTRHYFLAAVERLMDYGMSSPHYLRNR	1740
Db	1681	IDYDPTLSVEKKEDDPDIYEDENQSPRSFQKTRHYFLAAVERLMDYGMSSPHYLRNR	1740
Qy	1741	AQSGSVQPKFVYVQFETDGSFTQPLYRGEINHLGLDPTIRAEVENIIVTRNQAAS	1800
Db	1741	AQSGSVQPKFVYVQFETDGSFTQPLYRGEINHLGLDPTIRAEVENIIVTRNQAAS	1800
Qy	1801	PYSFYSSLISYEDDQAGAPKRNKFNKNETKTYEMKVOHMAPIKDEPCKANAYSDV	1860
Db	1801	PYSFYSSLISYEDDQAGAPKRNKFNKNETKTYEMKVOHMAPIKDEPCKANAYSDV	1860
Qy	1861	DEKDVHSGILGILVCHNTLNPAGHQVYVQEFALFTFIDETKSMYFTENNERNCRA	1920
Db	1861	DEKDVHSGILGILVCHNTLNPAGHQVYVQEFALFTFIDETKSMYFTENNERNCRA	1920
Qy	1921	PCNIOEDPTFKENTFRHALNGIIMDTLPGLYMAODRIRMTYLSGNSNENHSHIFSGH	1980
Db	1921	PCNIOEDPTFKENTFRHALNGIIMDTLPGLYMAODRIRMTYLSGNSNENHSHIFSGH	1980
Qy	1981	VFTVRKKEEYKALYNYLPGVEYVEMLPSTAGIWRVECLIGBHLHAGKSTLFLYSNKC	2040
Db	1981	VFTVRKKEEYKALYNYLPGVEYVEMLPSTAGIWRVECLIGBHLHAGKSTLFLYSNKC	2040
Qy	2041	QTPPLMASGHIRDFQITASGOYQONAPKILARLHSGSINAMSTKEPFSNIVDLAPMI	2100
Db	2041	QTPPLMASGHIRDFQITASGOYQONAPKILARLHSGSINAMSTKEPFSNIVDLAPMI	2100
Qy	2101	HGKTQAGAROKESSLYTSOPITMYSLDGKKNOTRGSTGLTAVFPGVNDSSGIRHNIFN	2160
Db	2101	HGKTQAGAROKESSLYTSOPITMYSLDGKKNOTRGSTGLTAVFPGVNDSSGIRHNIFN	2160
Qy	2161	PLIARIYRLHPHYSTRSTLRMELMGLDNLSCNMPGMSKAIISAOITVASSYFTNFA	2220
Db	2161	PLIARIYRLHPHYSTRSTLRMELMGLDNLSCNMPGMSKAIISAOITVASSYFTNFA	2220
Qy	2221	TWSPKARLHLQGRSNMARPQVNNPKEMLOVDQFKTKVYGVTTQGVKSLTSMYKREFL	2280
Db	2221	TWSPKARLHLQGRSNMARPQVNNPKEMLOVDQFKTKVYGVTTQGVKSLTSMYKREFL	2280
Qy	2281	ISSSDGHOVTLFQONKRVYVFGONDSTFPVNSLDPPLTFRYLRIHPQSVHQAIRLM	2340
Db	2281	ISSSDGHOVTLFQONKRVYVFGONDSTFPVNSLDPPLTFRYLRIHPQSVHQAIRLM	2340
Qy	2341	EVLGCEADQDY 2351	
Db	2341	EVLGCEADQDY 2351	
RESULT 14			
AAW11387 standard; Protein: 2351 AA.			
AAW11387;			
18-NOV-1997 (first entry)			
Active Factor VIII:C analogue N357X.			
Factor VIII:C; analogue; glycoprotein; blood coagulation cascade;			
fibrinogen; fibrin clot; haemostasis; haemophilia A; bleeding diathesis;			
plasma protease; thrombin; immunogen; antibody; haemophilia; therapy;			
proteolytic cleavage.			
Homo sapiens.			
Synthetic.			
Key			
Peptide			
Location/Qualifiers			
1..19			
/note="signal peptide"			
20..2351			
Protein			

FT		/note="mature Factor VIII:C"
FT	Region	20..1667
FT	Modified-site	/note="heavy chain fragment"
FT		376
FT	Region	/label-Phe, Glu, Pro
FT		1668..2350
FT	Domain	/note="light chain fragment"
FT		760..1667
FT		/note="B domain"
PN		MO9703195-A1.
XX		
PD	30-JAN-1997.	
XX		
PP	09-JUL-1996;	96MO-US11444.
XX		
PR	11-JUL-1995;	9505-0001025.
XX		
PA	(CHIR) CHIRON CORP.	
XX		
P1	Cohen FE, Hung DT, Innis M;	
XX		
DR	WPI; 1997-119050/11.	
XX		
PT	Factor VIII:C analog modified adjacent to a non-activating Arg	
PT	residue - used in the treatment of haemophiliacs, by improvement of	
PT	haemostasis	
XX		
PS	Claim 20; Page -; 90pp; English.	
XX		
CC	AA11330-W11472 represent active Factor VIII:C analogues of the	
CC	invention. These sequences were created by mutating the wild type Factor	
CC	VIII:C coding sequence (see AA751357) using mutagenic primers. The	
CC	analogues comprise a native Factor VIII:C polypeptide modified at a site	
CC	adjacent to a non-activating Arg residue so that a Arg-Pro or Pro-Arg	
CC	dipeptide is created. Factor VIII:C is a large glycoprotein that	
CC	participates in the blood coagulation cascade that ultimately converts	
CC	soluble fibrinogen to insoluble fibrin clot, effecting haemostasis. A	
CC	deficiency in Factor VIII:C is responsible for haemophilia A, which is an	
CC	X-chromosome-linked inherited bleeding diathesis. Factor VIII:C is	
CC	activated by plasma proteases, such as thrombin. During activation the	
CC	mature polypeptide is cleaved to generate heavy and light chain fragments	
CC	that are further cleaved. Complexes of two or more of the analogues,	
CC	nucleic acids and vectors encoding them may be used alone or in	
CC	conjunction with each other, for the prevention or treatment of active	
CC	Factor VIII:C deficiency in a mammal. The analogues may be used as	
CC	immunogens to raise antibodies, and in the treatment of haemophiliacs, by	
CC	improvement of haemostasis. The analogues are resistant to proteolytic	
CC	cleavage and display increased plasma half-life. They may be administered	
CC	at lower dosages and by different modes of administration.	
CC		
XX		
SO	Sequence 2351 AA;	
	Query Match 100.0%; Score 12413; DB 16; Length 2351;	
	Best Local Similarity 100.0%; Pred. No. 0;	
	Matches 2350; Conservative 0; Mismatches 1; Indels 0; Gaps 0;	

QY	241	AASARAPKMHVNGVYKRSPLGLGCHRSKYVWHVIGKGTPEVHSIFLEGHTFLVRNH	300
DB	241	AASARAPKMHVNGVYKRSPLGLGCHRSKYVWHVIGKGTPEVHSIFLEGHTFLVRNH	300
QY	301	QASLEISPTFTLTQATLMDGQFLPCHTSHOHQDEAVAYKDSCEPREPOLRKKNNE	360
DB	301	QASLEISPTFTLTQATLMDGQFLPCHTSHOHQDEAVAYKDSCEPREPOLRKKNNE	360
QY	361	EADYDDLTDSEMDVYREDDNSPSFTQIRSAVAKHPRPTVWHVIAAEEDMDYAPLVLA	420
DB	361	EADYDDLTDSEMDVYREDDNSPSFTQIRSAVAKHPRPTVWHVIAAEEDMDYAPLVLA	420
QY	421	PDRSYKSOYLNNGPORIGRKRYKAFNAATDEFTTRAIOHESGILGPLLYGEGDTL	480
DB	421	PDRSYKSOYLNNGPORIGRKRYKAFNAATDEFTTRAIOHESGILGPLLYGEGDTL	480
QY	481	LIFRNOASRPYIYHGTIDYRPLYSRRLPKGVKHLKQPLLPQELIKYKMYVYEDGP	540
DB	481	LIFRNOASRPYIYHGTIDYRPLYSRRLPKGVKHLKQPLLPQELIKYKMYVYEDGP	540
QY	541	TKSDPCLTRYSSFYNNMERDLASGLIGPLLCYKBSVDQRNQJMSDRNVILFSVDE	600
DB	541	TKSDPCLTRYSSFYNNMERDLASGLIGPLLCYKBSVDQRNQJMSDRNVILFSVDE	600
QY	601	NRSWYLFENIORFLPMPAGVQLEDEPEQASINHSINGVYFDSQLSCLEHVAWYIIS	660
DB	601	NRSWYLFENIORFLPMPAGVQLEDEPEQASINHSINGVYFDSQLSCLEHVAWYIIS	660
QY	661	IGAQTDLSVFESEGYFFKHKMYEETLTLPFSEGYFPMKPEPLMLGCHNSDFNRK	720
DB	661	IGAQTDLSVFESEGYFFKHKMYEETLTLPFSEGYFPMKPEPLMLGCHNSDFNRK	720
QY	721	MTALIKVSCDKNTGYEDSYEDISAYLLSKNNAIEPNSFSONSRAPSTROKOFNATYI	780
DB	721	MTALIKVSCDKNTGYEDSYEDISAYLLSKNNAIEPNSFSONSRAPSTROKOFNATYI	780
QY	781	PENDIEKTDMPFARHPMPKIONVSSDLMLTLOSPTPGLSLSDLOEAYEFTSDPS	840
DB	781	PENDIEKTDMPFARHPMPKIONVSSDLMLTLOSPTPGLSLSDLOEAYEFTSDPS	840
QY	841	PCADSNNSISEMTHRRPOLHSGMVTTPSSGQLNLNKKLTNNALDKKDFKVSST	900
DB	841	PCADSNNSISEMTHRRPOLHSGMVTTPSSGQLNLNKKLTNNALDKKDFKVSST	900
QY	901	SNMLISTIPSDNLAAGDTWSSLAGPSPVHYDSQDLTTLFGKSSPLTESGPLSSEE	960
DB	901	SNMLISTIPSDNLAAGDTWSSLAGPSPVHYDSQDLTTLFGKSSPLTESGPLSSEE	960
QY	961	NNDSKILESGLMSOESSMGKNVSTESGRFLFKGRAGPALLTDNMLFYVSIILKTN	1020
DB	961	NNDSKILESGLMSOESSMGKNVSTESGRFLFKGRAGPALLTDNMLFYVSIILKTN	1020
QY	1021	KTSNNSATNRKTHIDGPSLLIENSQVONILSDPTFEKKVTPLIHDMLMDKNAATLR	1080
DB	1021	KTSNNSATNRKTHIDGPSLLIENSQVONILSDPTFEKKVTPLIHDMLMDKNAATLR	1080
QY	1081	NHMSKNTTSSKNMEVQOKKEGPIPPDAQNDMSFFKMLFLPESARWIOPTHGNSLNSG	1140
DB	1081	NHMSKNTTSSKNMEVQOKKEGPIPPDAQNDMSFFKMLFLPESARWIOPTHGNSLNSG	1140
QY	1141	QGPSPKOLVSLGPEKSEVQONFLSEKNVYVVGKEEFLKDGKLEWFPSSRNLTMLDN	1200
DB	1141	QGPSPKOLVSLGPEKSEVQONFLSEKNVYVVGKEEFLKDGKLEWFPSSRNLTMLDN	1200
QY	1201	LHENNTNHOEKKIOELIEKKETLIDENYVLIQHTVGTNPKNKLFLISTROWVESYD	1260
DB	1201	LHENNTNHOEKKIOELIEKKETLIDENYVLIQHTVGTNPKNKLFLISTROWVESYD	1260
QY	1261	GAYAVLQDFRSLNDSTNRKTKTAHESKKEEENLGLGNQTKOIVKVACTTRISPT	1320
DB	1261	GAYAVLQDFRSLNDSTNRKTKTAHESKKEEENLGLGNQTKOIVKVACTTRISPT	1320
QY	1321	SOQNFVQSRKALKQFLPLEETELERRIIVDTSTQMSKNKHLTSTLTLOIDYNEKE	1380

Dd	1321	SOQNFVQSRKALFOFRIPLEETLEEKRIYIVDSTOWSKMMLHPLPSTLQIYNAKE	1380
Qy	1381	KGALTQSPLESDCLTRHSHIPQANSRPLPIKASPSPSIRPIYLRLVLFQONSHLPAASY	1440
Dd	1381	KGALTQSPLESDCLTRHSHIPQANSRPLPIKASPSPSIRPIYLRLVLFQONSHLPAASY	1440
Qy	1441	RKDSGOVSHSLQSAKKNMLSLATLLTLMGDQREVGSGTSATSYKRYKVENYLP	1500
Dd	1441	RKDSGOVSHSLQSAKKNMLSLATLLTLMGDQREVGSGTSATSYKRYKVENYLP	1500
Qy	1501	KPDLPTSGVELLPRVHVIYQKDLFPPTETISNSPGHLDIVEGSLTQSGALIKMEANRP	1560
Dd	1501	KPDLPTSGVELLPRVHVIYQKDLFPPTETISNSPGHLDIVEGSLTQSGALIKMEANRP	1560
Qy	1561	GKVPFLVNAEESAKPSPKLDPLAMONHNGTQIPREEMKSOEKSPPKTLFFKAKTITSL	1620
Dd	1561	GKVPFLVNAEESAKPSPKLDPLAMONHNGTQIPREEMKSOEKSPPKTLFFKAKTITSL	1620
Qy	1621	NACSNNAIAIMEGONKRELEVTMAQGTETLSCONPVLAKHNOEIRRTTISDOE	1680
Dd	1621	NACSNNAIAIMEGONKRELEVTMAQGTETLSCONPVLAKHNOEIRRTTISDOE	1680
Qy	1681	IDYDDTISYEMAKEDDITIDEDENQSRSQKTKRHTPIAAYEMLDYGSSSPHYLNR	1740
Dd	1681	IDYDDTISYEMAKEDDITIDEDENQSRSQKTKRHTPIAAYEMLDYGSSSPHYLNR	1740
Qy	1741	AOSGSVPQFKVVOEFHTDGSFTQPLYRGELNHELGLGPYIRAEVEDNITMVFERNQASR	1800
Dd	1741	AOSGSVPQFKVVOEFHTDGSFTQPLYRGELNHELGLGPYIRAEVEDNITMVFERNQASR	1800
Qy	1801	PYSEFSSILIEYEDDOGOAGRPKNFVNPNTKTYEKFVKOHNMAFPKDEFPCKAMAYESDV	1860
Dd	1801	PYSEFSSILIEYEDDOGOAGRPKNFVNPNTKTYEKFVKOHNMAFPKDEFPCKAMAYESDV	1860
Qy	1861	DLEKDVSHSGLGLVCHNTLNPAHQRQTVQOEFLLFTJFDETKSMYTEEMENCR	1920
Dd	1861	DLEKDVSHSGLGLVCHNTLNPAHQRQTVQOEFLLFTJFDETKSMYTEEMENCR	1920
Qy	1921	PONIONMDPFRKKNYFPAHNGYIMDTLPLVMAOORPRWYLLSMGSNENHSHIFGSH	1980
Dd	1921	PONIONMDPFRKKNYFPAHNGYIMDTLPLVMAOORPRWYLLSMGSNENHSHIFGSH	1980
Qy	1981	VFTYRKRKEEKALVYLIPGVFEVEMLPSKACIWRBECLIGEHLAAGNSTLEFVYSNKC	2040
Dd	1981	VFTYRKRKEEKALVYLIPGVFEVEMLPSKACIWRBECLIGEHLAAGNSTLEFVYSNKC	2040
Qy	2041	QTPPLGMSGHIRFOQTASAGQYOMAKRLRLHYSGSISIANSTKEPESWIKVDLAPMI	2100
Dd	2041	QTPPLGMSGHIRFOQTASAGQYOMAKRLRLHYSGSISIANSTKEPESWIKVDLAPMI	2100
Qy	2101	HGKTOGAKRKFSLTISOFIIMYSILDGKMOYRGRNSGTLMFPPANDNSGIKHNIFN	2160
Dd	2101	HGKTOGAKRKFSLTISOFIIMYSILDGKMOYRGRNSGTLMFPPANDNSGIKHNIFN	2160
Qy	2161	PPITARIYRIHAPHYSINSTLAMELMEGCDLNSCMFLGMEKASISDQITASSYTFNMA	2220
Dd	2161	PPITARIYRIHAPHYSINSTLAMELMEGCDLNSCMFLGMEKASISDQITASSYTFNMA	2220
Qy	2221	TWSPSKARHLHOGRSNAMPPOVNNPKEMLDVDFQKTMKATGYTITQGVKSLTSMYKEFL	2280
Dd	2221	TWSPSKARHLHOGRSNAMPPOVNNPKEMLDVDFQKTMKATGYTITQGVKSLTSMYKEFL	2280
Qy	2281	ISSSDQDGHOMTLFFQNGKRYKVOGQOSFPVYNSLADPLLTRTLRLRHPSWVHOALYRM	2340
Dd	2281	ISSSDQDGHOMTLFFQNGKRYKVOGQOSFPVYNSLADPLLTRTLRLRHPSWVHOALYRM	2340
Qy	2341	EVLGCEADOLY 2351	
Dd	2341	EVLGCEADOLY 2351	

AAW1362	
ID	AAW1362 standard; Protein; 2351 AA.
XX	
AC	AAW1362;
XX	
DT	18-NOV-1997 (first entry)
XX	
DE	Active Factor VIII:C analogue I277X.
XX	
KM	Factor VIII:C; analogue; glycoprotein; blood coagulation cascade;
KM	fibrinogen; fibrin clot; haemostasis; haemophilia A; bleeding diathesis;
KM	plasma protease; thrombin; immunogen; antibody; haemophilic; therapy;
XX	proteolytic cleavage.
XX	
OS	Homo sapiens.
OS	Synthetic.
XX	
FT	Key
FT	Location/Qualifiers
FT	Peptide
FT	1..19
FT	/note="signal peptide"
FT	20..2351
FT	/note="mature Factor VIII:C"
FT	20..1667
FT	/note="heavy chain fragment"
FT	296
FT	/label="Phe, Glu, Pro
FT	1668..2350
FT	/note="light chain fragment"
FT	760..1667
FT	/note="B domain"
XX	
PN	MO9703195-AA.
XX	
XX	30-JAN-1997.
XX	
PF	09-JUL-1996; 96MO-US11444.
XX	
PR	11-JUL-1995; 95US-0001025.
XX	
XX	(CHIR) CHIRON CORP.
PA	
PI	Cohen FE, Hung DT, Innis M;
XX	
XX	WPI: 1997-119050/11.
XX	
PT	Factor VIII:C analog modified adjacent to a non-activating Arg
PT	residue - used in the treatment of haemophiliacs, by improvement of
PT	haemostasis
XX	
PS	Claim 14, Page -: 90pp; English.
XX	
CC	AAW1330-W11472 represent active Factor VIII:C analogues of the
CC	invention. These sequences were created by mutating the wild type Factor
CC	VIII:C coding sequence (see AAT1357) using mutagenic primers. The
CC	analogues comprise a native Factor VIII:C polypeptide modified at a site
CC	adjacent to a non-activating Arg residue so that a Arg-Pro or Pro-Arg
CC	participates in the blood coagulation cascade that ultimately converts
CC	soluble fibrinogen to insoluble fibrin clot, effecting haemostasis. A
CC	deficiency in Factor VIII:C is responsible for haemophilia A, which is an
CC	X-chromosome-linked inherited bleeding diathesis. Factor VIII:C is
CC	activated by plasma proteases, such as thrombin. During activation the
CC	mature polypeptide is cleaved to generate heavy and light chain fragments
CC	that are further cleaved. Complexes of two or more of the analogues,
CC	nucleic acids and vectors encoding them may be used alone or in
CC	conjunction with each other, for the prevention or treatment of active
CC	Factor VIII:C deficiency in a mammal. The analogues may be used as
CC	immunogens to raise antibodies, and in the treatment of haemophiliacs, by
CC	improvement of haemostasis. The analogues are resistant to proteolytic
CC	cleavage and display increased plasma half-life. They may be administered
CC	at lower dosages and by different modes of administration.
XX	
Sequence	2351 AA;

Query Match 100.0%; Score 12413; DB 18; Length 2351;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2350; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 M0ELSTCFCLLRFCSATRRYIIGAVELSDMOSDLGELPYDARRPPRPSPFN 60
DB 1 M0ELSTCFCLLRFCSATRRYIIGAVELSDMOSDLGELPYDARRPPRPSPFN 60
QY 61 TSVYKRTLFVEFDHLENIAPRPMMGLPTIOAEVYDVVTYILKNASHPVSLAV 120
DB 61 TSVYKRTLFVEFDHLENIAPRPMMGLPTIOAEVYDVVTYILKNASHPVSLAV 120
QY 121 GSVYKASGAEYDQTSOREKEDKVPFGSGHTVYMOVLKENGPMASDPLCTYSYLH 180
DB 121 GSVYKASGAEYDQTSOREKEDKVPFGSGHTVYMOVLKENGPMASDPLCTYSYLH 180
QY 181 VDLKDNGLGALLVREGSLAKERTQTLKFTLLFVYDEKSHSTKNSLMORD 240
DB 181 VDLKDNGLGALLVREGSLAKERTQTLKFTLLFVYDEKSHSTKNSLMORD 240
QY 241 AASARAPMAMTVNGYVNRSLPGLIGCHRSYVMHYIGMTTPVHSIFLEGHTLVNRH 300
DB 241 AASARAPMAMTVNGYVNRSLPGLIGCHRSYVMHYIGMTTPVHSIFLEGHTLVNRH 300
QY 301 ROASLEISPTFLTAQTLMDLQGFLLFCHISSHODGMEAYVYVDSCEPEPQLRKKN 360
DB 301 ROASLEISPTFLTAQTLMDLQGFLLFCHISSHODGMEAYVYVDSCEPEPQLRKKN 360
QY 361 EADYDDDLTDESMVYVFPDDNSPFIQISYAKKHRTVHYIAAEEDMVAFLYA 420
DB 361 EADYDDDLTDESMVYVFPDDNSPFIQISYAKKHRTVHYIAAEEDMVAFLYA 420
QY 421 PDDRSYKSOYLNNGPQIRIGRRYKRVPMAYTDEFTKRBALQHSGLGLYXGVGDTL 480
DB 421 PDDRSYKSOYLNNGPQIRIGRRYKRVPMAYTDEFTKRBALQHSGLGLYXGVGDTL 480
QY 481 LIIFKNQASRPYNIYPHGITDVRPLYSRLLPKVKHLKDFLLFGEJFKKMTVYEDP 540
DB 481 LIIFKNQASRPYNIYPHGITDVRPLYSRLLPKVKHLKDFLLFGEJFKKMTVYEDP 540
QY 541 TRSDPCLTRYSSVYVNERDLASGLIGLLCYKESYVQDQNGOIMDKRNVILFVPE 600
DB 541 TRSDPCLTRYSSVYVNERDLASGLIGLLCYKESYVQDQNGOIMDKRNVILFVPE 600
QY 601 NRSWYLTENIORPLNPAGVQLEDPEOASNIMHSINGVYFDSIQLSVCLHEVAYYILS 660
DB 601 NRSWYLTENIORPLNPAGVQLEDPEOASNIMHSINGVYFDSIQLSVCLHEVAYYILS 660
QY 661 IGAQTDPLSVFSGYTFKHKMYEDTLTFPFSGETVEMSKENPGLMILGCHNSDFRNG 720
DB 661 IGAQTDPLSVFSGYTFKHKMYEDTLTFPFSGETVEMSKENPGLMILGCHNSDFRNG 720
QY 721 MTALLVSSCDKNTGDYEDSEDISAUYLSKNNALTEPSPNSRHSPTROKOFNATTI 780
DB 721 MTALLVSSCDKNTGDYEDSEDISAUYLSKNNALTEPSPNSRHSPTROKOFNATTI 780
QY 781 PENDIEKTDWFAHRTPMFKIIONVSSDDLMLLROSPPRHGSLSDLOEAKYETFDSDS 840
DB 781 PENDIEKTDWFAHRTPMFKIIONVSSDDLMLLROSPPRHGSLSDLOEAKYETFDSDS 840
QY 841 PGALDSNNSLSEKTEHPROLHBSGDMVTPESGLQLRLNEKIGTTAATELKLIDFVYST 900
DB 841 PGALDSNNSLSEKTEHPROLHBSGDMVTPESGLQLRLNEKIGTTAATELKLIDFVYST 900
QY 901 SNNLISTIPSDNLAACTDNTSSIGPSPMVHYISQDITTLFGKSSPLTSSGGPGLSLEE 960
DB 901 SNNLISTIPSDNLAACTDNTSSIGPSPMVHYISQDITTLFGKSSPLTSSGGPGLSLEE 960
QY 961 NNSDKLLESGLANSSOESSWGKRVSTESGRLFKGRABGAPALLTRKNNALFKVSIILKTN 1020
DB 961 NNSDKLLESGLANSSOESSWGKRVSTESGRLFKGRABGAPALLTRKNNALFKVSIILKTN 1020

QY 1021 KTSNNSATNRKTHIDGSLLIENSFSVMONLIESDTEFKKVTPLIHDRMLADKNATALRL 1080
DB 1021 KTSNNSATNRKTHIDGSLLIENSFSVMONLIESDTEFKKVTPLIHDRMLADKNATALRL 1080
QY 1081 NNSKNTSSKNNMVOOKKEEPIPPDQNDPMSPFIMLPPESAPVIOIRHGRKNSLNG 1140
DB 1081 NNSKNTSSKNNMVOOKKEEPIPPDQNDPMSPFIMLPPESAPVIOIRHGRKNSLNG 1140
QY 1141 QGSPKQVLSLPEKSVEGONFLSEKKNVYVKGFTKQDGLAEWVPSSRNLFTNLIN 1200
DB 1141 QGSPKQVLSLPEKSVEGONFLSEKKNVYVKGFTKQDGLAEWVPSSRNLFTNLIN 1200
QY 1201 LHENNTHNOEKKIOEIEKRETLIOENNVLPQIRHYVGTKNFKNLFLLSTRONVSSYD 1260
DB 1201 LHENNTHNOEKKIOEIEKRETLIOENNVLPQIRHYVGTKNFKNLFLLSTRONVSSYD 1260
QY 1261 GAYAPYLDPSLNDSTNRKNTAHFSKKGEBNLSGONKQIYEVYACTPISPT 1320
DB 1261 GAYAPYLDPSLNDSTNRKNTAHFSKKGEBNLSGONKQIYEVYACTPISPT 1320
QY 1321 SOONFYTORSKRALKORPLPLEETELKRIYVDTSTOWSKNNKHLTPSTLQIDYNEKE 1380
DB 1321 SOONFYTORSKRALKORPLPLEETELKRIYVDTSTOWSKNNKHLTPSTLQIDYNEKE 1380
QY 1381 KGAITQSPSLDCLTRSHSIPQANRSPPLIAKVSFPISIRPLYLRFVPODNSHLPASY 1440
DB 1381 KGAITQSPSLDCLTRSHSIPQANRSPPLIAKVSFPISIRPLYLRFVPODNSHLPASY 1440
QY 1441 RKDQSVOSSHFTOGAKNNLSIALITLTEMGDQREVGSIGTSATNSTYKRVNTLP 1500
DB 1441 RKDQSVOSSHFTOGAKNNLSIALITLTEMGDQREVGSIGTSATNSTYKRVNTLP 1500
QY 1501 KPDLPTSGKVELLPKHYIYOKDLPFTESNGSPGHLDLVBSLLOGTBCAIKMEANRP 1560
DB 1501 KPDLPTSGKVELLPKHYIYOKDLPFTESNGSPGHLDLVBSLLOGTBCAIKMEANRP 1560
QY 1561 GKVPFLRVATESAATPSKLLDPLANDNHYGQIQKEEMKSOEKSEPKTAFFKKKTJLSL 1620
DB 1561 GKVPFLRVATESAATPSKLLDPLANDNHYGQIQKEEMKSOEKSEPKTAFFKKKTJLSL 1620
QY 1621 NACESNHAIATINEONKPELEYENAKOGRTERLCSOPPLAKHOREITRTTLOSDEE 1680
DB 1621 NACESNHAIATINEONKPELEYENAKOGRTERLCSOPPLAKHOREITRTTLOSDEE 1680
QY 1681 IDYDDTISYEMKKEEDFDIYDENQSPRSPOKTRHYFLAVERLMDYGKSSPHYLRNR 1740
DB 1681 IDYDDTISYEMKKEEDFDIYDENQSPRSPOKTRHYFLAVERLMDYGKSSPHYLRNR 1740
QY 1741 AOGSVPOFKKVVFOEFTDGSFTOPLYRGLNEHGLGPIYRAVEYEDNIMYTFRNOASR 1800
DB 1741 AOGSVPOFKKVVFOEFTDGSFTOPLYRGLNEHGLGPIYRAVEYEDNIMYTFRNOASR 1800
QY 1801 PYSFVSLISYEEDROGAEPKRNKYKPNETKYVMKVOHMAPTKDEPCXAMAYFSDV 1860
DB 1801 PYSFVSLISYEEDROGAEPKRNKYKPNETKYVMKVOHMAPTKDEPCXAMAYFSDV 1860
QY 1861 DLEKDVHSGILGPLLYCHTNTLNPABROVYVOEALPFTTFDITKSWYTEMENRCRA 1920
DB 1861 DLEKDVHSGILGPLLYCHTNTLNPABROVYVOEALPFTTFDITKSWYTEMENRCRA 1920
QY 1921 PCNIOHEDPTFKENRFAFIINGYIMOTLGLVMAQDQIRMYLSMKSNNINISIFSGH 1980
DB 1921 PCNIOHEDPTFKENRFAFIINGYIMOTLGLVMAQDQIRMYLSMKSNNINISIFSGH 1980
QY 1981 VFTVRRKEEYKALYNLYRGVEFYVEMLSKAGIRVYECILGELHAGNSTFLVYYSKNC 2040
DB 1981 VFTVRRKEEYKALYNLYRGVEFYVEMLSKAGIRVYECILGELHAGNSTFLVYYSKNC 2040
QY 2041 QPPLGMAASHIRDFQITASGOYQMAFKLARLYSGSINANSTKEPSWIKYDLAPMI 2100
DB 2041 QPPLGMAASHIRDFQITASGOYQMAFKLARLYSGSINANSTKEPSWIKYDLAPMI 2100
QY 2101 HGKGTGAGRKSSLYISQFTIMSIDGKMKQTYRGNSTGTLMVFGNDVSSGIKHNIFN 2160

Db	2101	HGIKTQGAOKFSSLYISOFILMSIDGKKQTYRGNSTGTLAMFPGNVDSGSKHNIFN	2160
Qy	2161	PLIARYIRLHPTHTYSIRSTLRMLNGCDLNSCSMPLGMEKSAISDAQITASSYFTNMA	2220
Db	2161	PLIARYIRLHPTHTYSIRSTLRMLNGCDLNSCSMPLGMEKSAISDAQITASSYFTNMA	2220
Qy	2221	TWSPSKARLHLOGRBNAMRPQVNNPKEMLOVDFOKTMKVGTTOGYKSLTSMYKEFL	2280
Db	2221	TWSPSKARLHLOGRBNAMRPQVNNPKEMLOVDFOKTMKVGTTOGYKSLTSMYKEFL	2280
Qy	2281	ISSSQDGHQWTLFPONGKVKVTFQGNODSFTPPVNSLDPPLLRYLRIHPQSWHOIALRM	2340
Db	2281	ISSSQDGHQWTLFPONGKVKVTFQGNODSFTPPVNSLDPPLLRYLRIHPQSWHOIALRM	2340
Qy	2341	EVLGCEADPLY 2351	
Db	2341	EVLGCEADPLY 2351	

Search completed: July 2, 2003, 09:42:28
 Job time : 111 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2003, 09:40:36 ; Search time 33 Seconds
(without alignments)
2096.160 Million cell updates/sec

Title: NP_000123

Sequence: 1 mgetstcfcilclrfcfa.....vwhjalrmeivceagdy 2351

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Issued Patents, AA.*
1: /cgn2_6/ptodata/1/laa/5A.COMB.pep.*
2: /cgn2_6/ptodata/1/laa/5B.COMB.pep.*
3: /cgn2_6/ptodata/1/laa/5A.COMB.pep.*
4: /cgn2_6/ptodata/1/laa/5B.COMB.pep.*
5: /cgn2_6/ptodata/1/laa/PCRM.COMB.pep.*
6: /cgn2_6/ptodata/1/laa/Backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	12412	100.0	2351	6	5171844-2	Patent No. 5171844
2	12409	99.9	2351	1	US-08-366-851A-2	Sequence 2, Appl1
3	12405	99.9	2351	1	US-08-121-202-2	Sequence 2, Appl1
4	12396	99.8	2351	6	5422260-1	Patent No. 5422260
5	12301	99.1	2332	1	US-07-864-004B-4	Sequence 4, Appl1
6	12301	99.1	2332	1	US-08-251-937A-4	Sequence 2, Appl1
7	12301	99.1	2332	1	US-08-212-133A-2	Sequence 2, Appl1
8	12301	99.1	2332	1	US-08-474-503-2	Sequence 2, Appl1
9	12301	99.1	2332	2	US-08-670-707A-2	Sequence 2, Appl1
10	12301	99.1	2332	4	US-09-037-601-2	Sequence 2, Appl1
11	12301	99.1	2332	4	US-09-315-179-2	Sequence 2, Appl1
12	12301	99.1	2332	4	US-09-523-656-2	Sequence 2, Appl1
13	12301	99.1	2332	5	PCT-US93-02275-4	Sequence 2, Appl1
14	12301	99.1	2332	5	PCT-US94-13200-2	Sequence 2, Appl1
15	12294	99.0	2332	1	US-08-276-594A-2	Sequence 2, Appl1
16	12251	98.7	2332	4	US-09-324-867-3	Sequence 3, Appl1
17	9493	76.4	2343	4	US-09-324-867-3	Sequence 3, Appl1
18	8835	71.1	2319	1	US-08-212-133A-8	Sequence 8, Appl1
19	8835	71.1	2319	1	US-08-474-503-6	Sequence 6, Appl1
20	8835	71.1	2319	2	US-08-670-707A-6	Sequence 6, Appl1
21	8835	71.1	2319	4	US-09-037-601-6	Sequence 6, Appl1
22	8835	71.1	2319	4	US-09-315-179-6	Sequence 6, Appl1
23	8835	71.1	2319	4	US-09-523-656-28	Sequence 28, Appl1
24	8730	70.3	2304	5	PCT-US94-13200-6	Sequence 6, Appl1
25	8730	70.3	2304	4	US-09-324-867-4	Sequence 4, Appl1
26	8357	67.3	1661	2	US-08-882-083-2	Sequence 2, Appl1
27	8357	67.3	1661	2	US-08-558-107-2	Sequence 2, Appl1

28	8357	67.3	1661	4	US-09-243-539-2	Sequence 2, Appl1
29	8076	65.0	2133	2	US-08-670-707A-37	Sequence 37, Appl1
30	8076	65.0	2133	4	US-09-037-601-37	Sequence 37, Appl1
31	8076	65.0	2133	4	US-09-315-179-37	Sequence 37, Appl1
32	8070	65.0	2133	4	US-09-523-656-30	Sequence 30, Appl1
33	7959.5	64.1	2115	4	US-09-324-867-5	Sequence 5, Appl1
34	7400	59.6	1471	1	US-08-683-839B-3	Sequence 3, Appl1
35	7227	58.2	1438	4	US-09-209-916-1	Sequence 1, Appl1
36	6224	50.1	1467	4	US-09-523-656-38	Sequence 38, Appl1
37	6122	49.3	1443	4	US-08-670-707A-39	Sequence 39, Appl1
38	6122	49.3	1443	4	US-09-037-601-39	Sequence 39, Appl1
39	6122	49.3	1443	4	US-09-315-179-39	Sequence 39, Appl1
40	2768.5	22.3	2183	3	US-08-746-111-5	Sequence 3, Appl1
41	2444.5	15.7	541	1	US-08-121-202-4	Sequence 5, Appl1
42	2258	18.2	439	3	US-08-448-722A-5	Sequence 5, Appl1
43	2258	18.2	439	4	US-08-189-309B-5	Sequence 5, Appl1
44	2190.5	17.6	868	1	US-07-864-004B-6	Sequence 6, Appl1
45	2190.5	17.6	868	1	US-08-251-937A-6	Sequence 6, Appl1

ALIGNMENTS

RESULT 1
5171844-2
; Patent No. 5171844
; APPLICANT: VAN COEYEN, ALBERT J.J.; PANNENKOEK, HANS; VERBET,
; MARTINUS P.; VAN LEESEN, ROBERT W.
; TITLE OF INVENTION: PROTEIN WITH FACTOR VIII ACTIVITY
; PROCESS FOR THEIR PREPARATION USING GENETICALLY-ENGINEERED CELLS
; AND PHARMACEUTICAL COMPOSITIONS CONTAINING THEM
; NUMBER OF SEQUENCES: 12
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US-07/205,226
; FILING DATE: 10-JUN-1988
; SEQ ID NO: 2:
; LENGTH: 2351
5171844-2

Query Match 100.0%; Score 12412; DB 6; Length 2351;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2350; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY	1	MOTELSTCFCLLRFCFSATRRYYIGAVELSDMOYSDGLPELPVAPRPPVPSPPFN	60
DB	1	MOTELSTCFCLLRFCFSATRRYYIGAVELSDMOYSDGLPELPVAPRPPVPSPPFN	60
QY	61	TSVYVKKTLFEVFDHDLFNIAKPRPMGLGPTIOAEYDVTYITLKNMASHPVSLHAY	120
DB	61	TSVYVKKTLFEVFDHDLFNIAKPRPMGLGPTIOAEYDVTYITLKNMASHPVSLHAY	120
QY	121	GVSYWKAEGAEYDQTSOREKEDDKVFPGSGHYVVOYLKENGPMASDPLCLTYSTLSH	180
DB	121	GVSYWKAEGAEYDQTSOREKEDDKVFPGSGHYVVOYLKENGPMASDPLCLTYSTLSH	180
QY	181	VDLYMDNSGLIGALVCRGSLAKKKTOTLAKFLFAYPDGSKWHSZETNSLMODRD	240
DB	181	VDLYMDNSGLIGALVCRGSLAKKKTOTLAKFLFAYPDGSKWHSZETNSLMODRD	240
QY	241	AASARAPMPKRVNGVYVNSLPGLISCHRSYVHVIGMGTPEVHSIFLESGFTLVANH	300
DB	241	AASARAPMPKRVNGVYVNSLPGLISCHRSYVHVIGMGTPEVHSIFLESGFTLVANH	300
QY	301	QASLEISPIITFLNAOTLMDLGOPLIFCHISSHOHDMAYKRVVSCPEEQOLMKKNE	360
DB	301	QASLEISPIITFLNAOTLMDLGOPLIFCHISSHOHDMAYKRVVSCPEEQOLMKKNE	360
QY	361	EADYVDDDLTSEKDYVRFDDNSPSEIOIRSVAKKPKTMVHTAAEEDMDVAPLVTA	420
DB	361	EADYVDDDLTSEKDYVRFDDNSPSEIOIRSVAKKPKTMVHTAAEEDMDVAPLVTA	420
QY	421	PDRSKSYLANPKQIRKRYKRVAPVYDTEFKTRREALIOESSILPPLLYGEGDTL	480
DB	421	PDRSKSYLANPKQIRKRYKRVAPVYDTEFKTRREALIOESSILPPLLYGEGDTL	480

```

Db      421 PDDRSYKSOYLNNQPORIGRKYKVFMAAYTDEFTREAIQHESGILGPLLGEVGDIL 480
Qy      481 LIIFFKNOASRPYNIYPIGITTDRPLYSRRLPKYKHLKDPILIGPIFYKMTVYEDGP 540
Db      481 LIIFFKNOASRPYNIYPIGITTDRPLYSRRLPKYKHLKDPILIGPIFYKMTVYEDGP 540
Qy      541 TKSDFRCILTRYISFVWNERDLASGLIGPLLICYKESVDORGNOIMSKRNYILEFVDE 600
Db      541 TKSDFRCILTRYISFVWNERDLASGLIGPLLICYKESVDORGNOIMSKRNYILEFVDE 600
Qy      601 NRSWLTENIORFLRPNAGVLEDEPEFOASNIMHSINGYVDSILQSLVCLHEVAWTLLS 660
Db      601 NRSWLTENIORFLRPNAGVLEDEPEFOASNIMHSINGYVDSILQSLVCLHEVAWTLLS 660
Qy      661 IGAOFDPLSVFSGYFPHKMYEDLTLFPPSGEYFVFNEMRPGMLIGCHNSDFRRNG 720
Db      661 IGAOFDPLSVFSGYFPHKMYEDLTLFPPSGEYFVFNEMRPGMLIGCHNSDFRRNG 720
Qy      721 MTRLLKVSCKNTGDIYEDSYEDISATYLLSKNNALIEPFSFSONSRHSTROKOPNATTI 780
Db      721 MTRLLKVSCKNTGDIYEDSYEDISATYLLSKNNALIEPFSFSONSRHSTROKOPNATTI 780
Qy      781 PENDIEKTDPMFAHRTMPKIQNVSSDILMLAQSP7PHGISLSDIQAKYETESDDPS 840
Db      781 PENDIEKTDPMFAHRTMPKIQNVSSDILMLAQSP7PHGISLSDIQAKYETESDDPS 840
Qy      841 PGALDSNNSLSEMTNHFAPQLHSGDMVFTPESSGLOLRNEKLTGTATATLKLIDFKVYST 900
Db      841 PGALDSNNSLSEMTNHFAPQLHSGDMVFTPESSGLOLRNEKLTGTATATLKLIDFKVYST 900
Qy      901 SNNLITSTPNDNLAAQDNTSSIGSPSMVHYHDSQLODTLFGKSSPLRESGSPISLSEE 960
Db      901 SNNLITSTPNDNLAAQDNTSSIGSPSMVHYHDSQLODTLFGKSSPLRESGSPISLSEE 960
Qy      961 NNDKSLLESGIAMSQESSWCKNVSTESGRIFKGRAGHAPALLTKONALFKYSISILKTN 1020
Db      961 NNDKSLLESGIAMSQESSWCKNVSTESGRIFKGRAGHAPALLTKONALFKYSISILKTN 1020
Qy      1021 KTSNNSATNRKTHIDGSLIENSPSWONLIESDTEFKKVTPLIDHRMLDKNKTATLRL 1080
Db      1021 KTSNNSATNRKTHIDGSLIENSPSWONLIESDTEFKKVTPLIDHRMLDKNKTATLRL 1080
Qy      1081 NMSNKTSSSKNMWMOOKKEGPIPPDAONPMSPFKMLFLPSANMIORTGKKSLSNG 1140
Db      1081 NMSNKTSSSKNMWMOOKKEGPIPPDAONPMSPFKMLFLPSANMIORTGKKSLSNG 1140
Qy      1141 QGSPKQVLSIGPEKSYEGONFLSEKNKVVYVKGGEFTYDGLKEMVFPSSRLFLTINDN 1200
Db      1141 QGSPKQVLSIGPEKSYEGONFLSEKNKVVYVKGGEFTYDGLKEMVFPSSRLFLTINDN 1200
Qy      1201 LHENNTNHOEKKIOEIEKRETLIOENVVLPOIHRVTGKNFKMKNLFLSTRQONVEGSYD 1260
Db      1201 LHENNTNHOEKKIOEIEKRETLIOENVVLPOIHRVTGKNFKMKNLFLSTRQONVEGSYD 1260
Qy      1261 GAYAPVLODFRSLNDSTNRTKHTAHFSKKGSEENLEGLQOTKOIYEKACACTRISBPT 1320
Db      1261 GAYAPVLODFRSLNDSTNRTKHTAHFSKKGSEENLEGLQOTKOIYEKACACTRISBPT 1320
Qy      1321 SOONFYTQSKKALKQORLPLEETLEKRIYDOSTWQSKNMAHL7PSTLQOINDNEKE 1380
Db      1321 SOONFYTQSKKALKQORLPLEETLEKRIYDOSTWQSKNMAHL7PSTLQOINDNEKE 1380
Qy      1381 KGAITOSPISDCLTRSHSIPQANRSPPLAKVSSPSPSIRPIYTLRVLFQONSHSLPAASY 1440
Db      1381 KGAITOSPISDCLTRSHSIPQANRSPPLAKVSSPSPSIRPIYTLRVLFQONSHSLPAASY 1440
Qy      1441 RKKDSGVQSSSHFLOGAKNNLSIALITLTMGDORREVSIGTSATNSVYTKKVENTYVLP 1500
Db      1441 RKKDSGVQSSSHFLOGAKNNLSIALITLTMGDORREVSIGTSATNSVYTKKVENTYVLP 1500
Qy      1501 KPDLPTSCKVELLPKVAIYQKDLFPTESVNGSPGHDLVBSGLLOGTREGALKNWNEANRP 1560
Db      1501 KPDLPTSCKVELLPKVAIYQKDLFPTESVNGSPGHDLVBSGLLOGTREGALKNWNEANRP 1560

```

```

Qy      1561 GKVPELRAVATESAKTPSKLIDPLAMDNRHGTQIPKEBKNQSOEKSPERTAKKRODTLSTL 1620
Db      1561 GKVPELRAVATESAKTPSKLIDPLAMDNRHGTQIPKEBKNQSOEKSPERTAKKRODTLSTL 1620
Qy      1621 NACSENHAIAMINEGONKPEIEVMAKGRERELCSQNPVYKRRHORETRTQOSPOEE 1680
Db      1621 NACSENHAIAMINEGONKPEIEVMAKGRERELCSQNPVYKRRHORETRTQOSPOEE 1680
Qy      1681 IDYDFTISVEAKKEDFOIYDEBENOSPPSFOKTRHIFTAAVERLMDYGMSSPHYLRNR 1740
Db      1681 IDYDFTISVEAKKEDFOIYDEBENOSPPSFOKTRHIFTAAVERLMDYGMSSPHYLRNR 1740
Qy      1741 AQSQSVPOFKKVVFOEFTDGSFTOPLYNGELNETHLGLIPYIRAEVEDNIWYTERNOASR 1800
Db      1741 AQSQSVPOFKKVVFOEFTDGSFTOPLYNGELNETHLGLIPYIRAEVEDNIWYTERNOASR 1800
Qy      1801 PYSFYSLSIYSEEDQOGAEPKRNKVFKNETKTYFVKVYOHMAPTKDEFDCKAAYPSDV 1860
Db      1801 PYSFYSLSIYSEEDQOGAEPKRNKVFKNETKTYFVKVYOHMAPTKDEFDCKAAYPSDV 1860
Qy      1861 DLEKDVHSGILGPLYCHNTLNAPAGQVTVQEFALFTIYDETRKSYFENNERNCRA 1920
Db      1861 DLEKDVHSGILGPLYCHNTLNAPAGQVTVQEFALFTIYDETRKSYFENNERNCRA 1920
Qy      1921 PCNIOMEDPTEKRENTREHAINGYIMDTLPGLVMAQDQRIKWYLLSGSNENIHSIHSGH 1980
Db      1921 PCNIOMEDPTEKRENTREHAINGYIMDTLPGLVMAQDQRIKWYLLSGSNENIHSIHSGH 1980
Qy      1981 VFTYRKKEEYKALYNLYNPGVEFYEYEMLPKSAQIMWRECIJGEHLHAGSTLFLVYSNK 2040
Db      1981 VFTYRKKEEYKALYNLYNPGVEFYEYEMLPKSAQIMWRECIJGEHLHAGSTLFLVYSNK 2040
Qy      2041 QTPILMASGCHTRDQITASQOYQONAPPLAHLHSGSINAMSKKEPESMIVDLAPMT 2100
Db      2041 QTPILMASGCHTRDQITASQOYQONAPPLAHLHSGSINAMSKKEPESMIVDLAPMT 2100
Qy      2101 HGKTOGARQKSFSSLYISQFIIMYSIDSKKQIYRGNSGTGLWVFGVDSGIRKNIEN 2160
Db      2101 HGKTOGARQKSFSSLYISQFIIMYSIDSKKQIYRGNSGTGLWVFGVDSGIRKNIEN 2160
Qy      2161 PPIIARYIRLPHYISIRSTLRMELMGCDLNSCAMPJGMSKASISDAQTASSYFTNMA 2220
Db      2161 PPIIARYIRLPHYISIRSTLRMELMGCDLNSCAMPJGMSKASISDAQTASSYFTNMA 2220
Qy      2221 TWSPKARLHLOGSNAMRPOYNNPREMLQVDPQTKMYGVYTGQSVSLTSMYVEFL 2280
Db      2221 TWSPKARLHLOGSNAMRPOYNNPREMLQVDPQTKMYGVYTGQSVSLTSMYVEFL 2280
Qy      2281 ISSDQGHQWTLFQNGKAVTVQGNDSFTPVVNSLDPPLITRYLRIHQSVWQIALRM 2340
Db      2281 ISSDQGHQWTLFQNGKAVTVQGNDSFTPVVNSLDPPLITRYLRIHQSVWQIALRM 2340
Qy      2341 EYLGCEADOLY 2351
Db      2341 EYLGCEADOLY 2351

```

```

RESULT 2
US-08-366-851A-2
; Sequence 2, Application US/08366851A
; Patent No. 5681746
; GENERAL INFORMATION:
; APPLICANT: Bodner, Mordechai
; APPLICANT: De Polo, Nicolas J.
; APPLICANT: Hsu, David Chl-Tang
; APPLICANT: Chang, Steven
; TITLE OF INVENTION: Retroviral Delivery of Full Length Factor VIII
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Viagenne, Inc.
; STREET: 11055 Roselle Street
; CITY: San Diego

```


STATE: California
COUNTRY: U.S.A.
ZIP: 92121
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/366,851A
FILING DATE:
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Chambers, Daniel M.
REGISTRATION NUMBER: 34,561
REFERENCE/DOCKET NUMBER: 930049, 438
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 452-1288
TELEFAX: (619) 452-2616
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2351 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-366-851A-2

Query Match 99.9%; Score 12409; DB 1; Length 2351;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2350; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MOTELSTCFCLLRFCEFSATRRYYGAVELSDMDQSDGLDELVDARPPRPVPSFPPN 60
DB 1 MOTELSTCFCLLRFCEFSATRRYYGAVELSDMDQSDGLDELVDARPPRPVPSFPPN 60
QY 61 TSVYAKTLFEVFDHLFNLKPRPMMGLPTIOAEVYDTVYITLKNASHPSILNAV 120
DB 61 TSVYAKTLFEVFDHLFNLKPRPMMGLPTIOAEVYDTVYITLKNASHPSILNAV 120
QY 121 GVSVMASGAGAYDQDSQREKEDDYFPGSGSHYVQVLEKNGPMASDPLCTYSYLSH 180
DB 121 GVSVMASGAGAYDQDSQREKEDDYFPGSGSHYVQVLEKNGPMASDPLCTYSYLSH 180
QY 181 VDLVNDNSGLIGALLVCREGSLAKEKTYTLKFTLLFAVFDDECKSWHSETKNSLMQDD 240
DB 181 VDLVNDNSGLIGALLVCREGSLAKEKTYTLKFTLLFAVFDDECKSWHSETKNSLMQDD 240
QY 241 AASARAPKMHVYNGVYRSLPGLIGCHRSYVMHYIGMGTTPEVHSIFLEGHTFLVRH 300
DB 241 AASARAPKMHVYNGVYRSLPGLIGCHRSYVMHYIGMGTTPEVHSIFLEGHTFLVRH 300
QY 301 ROASLEASPTLTLKOTLLMDLQGLFLFCHISSHODGMEAYVYVDSCPPEPOLRKKNNE 360
DB 301 ROASLEASPTLTLKOTLLMDLQGLFLFCHISSHODGMEAYVYVDSCPPEPOLRKKNNE 360
QY 361 EADYDDDLTDEMDVYFDDNSPFIOTIRSAKKHPTVWYIAAEEDMDYAPLVLA 420
DB 361 EADYDDDLTDEMDVYFDDNSPFIOTIRSAKKHPTVWYIAAEEDMDYAPLVLA 420
QY 421 PDDRSYKSOYLNGFORIGRKYKVRFMAYTDEFTFTRALIOHSGILGGLTGEVDTL 480
DB 421 PDDRSYKSOYLNGFORIGRKYKVRFMAYTDEFTFTRALIOHSGILGGLTGEVDTL 480
QY 481 LIIFKNOASPNYNYPHGITVYRPLYSRLPKGVKHLADPFLTGELEFKTKMYVEDGP 540
DB 481 LIIFKNOASPNYNYPHGITVYRPLYSRLPKGVKHLADPFLTGELEFKTKMYVEDGP 540
QY 541 TKSDPRLCTRYSSFYNNERDLASGLIGPLLTICKEVSDQROQJMSDKRNVILFVDE 600
DB 541 TKSDPRLCTRYSSFYNNERDLASGLIGPLLTICKEVSDQROQJMSDKRNVILFVDE 600
QY 601 NRSWYLTENIQRFLENPAVGQLEDPEPQASNTMHSINGYVDSQLSCLHEVAAYTILS 660
DB 601 NRSWYLTENIQRFLENPAVGQLEDPEPQASNTMHSINGYVDSQLSCLHEVAAYTILS 660

DB 601 NRSWYLTENIQRFLENPAVGQLEDPEPQASNTMHSINGYVDSQLSCLHEVAAYTILS 660
QY 661 IGAOTDFLSVFFSGYTFKKKVVYEDTLTLPFSGETVMSNENGLIILCANSDFRNG 720
DB 661 IGAOTDFLSVFFSGYTFKKKVVYEDTLTLPFSGETVMSNENGLIILCANSDFRNG 720
QY 721 MTLALLVSSCDKNTDYYEDSDYEDISAYLLSKNNMIEPRSSQSSRRPSTROKOFNATTI 780
DB 721 MTLALLVSSCDKNTDYYEDSDYEDISAYLLSKNNMIEPRSSQSSRRPSTROKOFNATTI 780
QY 781 PENDIKTDPWFHFTPMKTIQNVSSDLMLKROSPPHGLSLSDLOEAKYTFESDDPS 840
DB 781 PENDIKTDPWFHFTPMKTIQNVSSDLMLKROSPPHGLSLSDLOEAKYTFESDDPS 840
QY 841 PGADISNNLSLMTFTRPOLHSGDMVTFPESGLQRLNEKLTGTTAATELKLDFEVSST 900
DB 841 PGADISNNLSLMTFTRPOLHSGDMVTFPESGLQRLNEKLTGTTAATELKLDFEVSST 900
QY 901 SNNLITSTSDMLAGTDTNTSSLAGPSPMPVHDSDLTTLTGKSSPULTSSGSPLSSE 960
DB 901 SNNLITSTSDMLAGTDTNTSSLAGPSPMPVHDSDLTTLTGKSSPULTSSGSPLSSE 960
QY 961 NNDKLLSEGLNNSQESSMGKNVSTESGRLFKGRAGAPALTLTKDNALFKVVISLKTN 1020
DB 961 NNDKLLSEGLNNSQESSMGKNVSTESGRLFKGRAGAPALTLTKDNALFKVVISLKTN 1020
QY 1021 KTSNNSATNRKTHIDGPLLLENSPVQMTILESDEFEKKTYPILHDMMDKNTATLRL 1080
DB 1021 KTSNNSATNRKTHIDGPLLLENSPVQMTILESDEFEKKTYPILHDMMDKNTATLRL 1080
QY 1081 NMSKNTSSKKNMEVQOKKEGPIPPPAQNPDMSPFKMLFPESARNIORTKRNLSNG 1140
DB 1081 NMSKNTSSKKNMEVQOKKEGPIPPPAQNPDMSPFKMLFPESARNIORTKRNLSNG 1140
QY 1141 QGSPKPOLVSLGPEKSVBQNFLEKKNVYVKGFTYDGLKEMVPESSRNILFTNLDN 1200
DB 1141 QGSPKPOLVSLGPEKSVBQNFLEKKNVYVKGFTYDGLKEMVPESSRNILFTNLDN 1200
QY 1201 LHEKNTHNOEKIIOEIEKKEFTLQENVVLPOLHYTGKKNMKNLFLSTRONVGSVD 1260
DB 1201 LHEKNTHNOEKIIOEIEKKEFTLQENVVLPOLHYTGKKNMKNLFLSTRONVGSVD 1260
QY 1261 GAYAPVLODFRSLNDSTNRKTHAFSKKGEENIEGLAGNOTKOIYEKTACTTRISPT 1320
DB 1261 GAYAPVLODFRSLNDSTNRKTHAFSKKGEENIEGLAGNOTKOIYEKTACTTRISPT 1320
QY 1321 SQQNFYQSRKALAKQFLPLEETELKRIIVDDSTQSKNMKHLTPSTLTQIDYNEKE 1380
DB 1321 SQQNFYQSRKALAKQFLPLEETELKRIIVDDSTQSKNMKHLTPSTLTQIDYNEKE 1380
QY 1381 KGATQSPISDCLTSHSITPOANNSPPIAKVSSPSTIRPIYTLRVLPONSSHPLAASY 1440
DB 1381 KGATQSPISDCLTSHSITPOANNSPPIAKVSSPSTIRPIYTLRVLPONSSHPLAASY 1440
QY 1441 RKKDSGVQESSHFLQAKKNNLSLAITLLEMTGDQREVSIGTSATNSVYKKRENTVLP 1500
DB 1441 RKKDSGVQESSHFLQAKKNNLSLAITLLEMTGDQREVSIGTSATNSVYKKRENTVLP 1500
QY 1501 KPDLPTSGKVELLPRVHIYQKDLFPTETSGNSPGHLDVBSGLQEGEAKIMNANRP 1560
DB 1501 KPDLPTSGKVELLPRVHIYQKDLFPTETSGNSPGHLDVBSGLQEGEAKIMNANRP 1560
QY 1561 GKVPFLRVATESAKTPSLDPLAMNNHGTQIPIREBKSOEKSPETVAFKKKDTLSL 1620
DB 1561 GKVPFLRVATESAKTPSLDPLAMNNHGTQIPIREBKSOEKSPETVAFKKKDTLSL 1620
QY 1621 NACSNHAIALINEGONKPELEVYMAKOGTERLCSQNPVLKRHOBEITRTTLQSDQEE 1680
DB 1621 NACSNHAIALINEGONKPELEVYMAKOGTERLCSQNPVLKRHOBEITRTTLQSDQEE 1680
QY 1681 IDYDDTISVEMKKEPDIYDDDENQSRSPQKTRHYFTAAVERLMQYGMSSPHYLARR 1740
DB 1681 IDYDDTISVEMKKEPDIYDDDENQSRSPQKTRHYFTAAVERLMQYGMSSPHYLARR 1740

QY 1741 AOSGVPQFKVYQFETDGSFTQPLNGELNEHLGLSPYIRAEVEDINIMTFRQASR 1800
DB 1741 AOSGVPQFKVYQFETDGSFTQPLNGELNEHLGLSPYIRAEVEDINIMTFRQASR 1800
QY 1801 PYSTYSSLISEEDQOQGEPRKNVKNENKTYFMKVQHNAPTKDEPDCKAANAYSKV 1860
DB 1801 PYSTYSSLISEEDQOQGEPRKNVKNENKTYFMKVQHNAPTKDEPDCKAANAYSKV 1860
QY 1861 DLEKDVHSGILGPLVCHNTINPAHQVQVQEFALFTIEDTKSMYFTEENMRNCRA 1920
DB 1861 DLEKDVHSGILGPLVCHNTINPAHQVQVQEFALFTIEDTKSMYFTEENMRNCRA 1920
QY 1921 PCNTOMEDPTFKENYFPHAINGYIMDTLPGVMAODORIRWYLLSGMSENHSHSGH 1980
DB 1921 PCNTOMEDPTFKENYFPHAINGYIMDTLPGVMAODORIRWYLLSGMSENHSHSGH 1980
QY 1981 VETVKKKEEKYKALYMLYGPVEYEMLPKAGIMRVCECLIGELHAGNSTLFLVYSNK 2040
DB 1981 VETVKKKEEKYKALYMLYGPVEYEMLPKAGIMRVCECLIGELHAGNSTLFLVYSNK 2040
QY 2041 QTPIGMASGHTRDFOITASGOYGNAPKLARLHYSGISNASTKEPFSWIKYDLAPMI 2100
DB 2041 QTPIGMASGHTRDFOITASGOYGNAPKLARLHYSGISNASTKEPFSWIKYDLAPMI 2100
QY 2101 HGKTQAGAROKFSSLYISQFIIMYSIDGKKWQYRGNSTGTLMVFFGVNDSSGINKHIFN 2160
DB 2101 HGKTQAGAROKFSSLYISQFIIMYSIDGKKWQYRGNSTGTLMVFFGVNDSSGINKHIFN 2160
QY 2161 PRITARYRLPHYYSRSTLRMEMLGCDLNSCMPGAMESKASDAQITASSYFTNMA 2220
DB 2161 PRITARYRLPHYYSRSTLRMEMLGCDLNSCMPGAMESKASDAQITASSYFTNMA 2220
QY 2221 TWSPSKARLHOGSNAMRPQVNNREMLQVDFQTKMAYGVTTQYKSLISMTYKEFL 2280
DB 2221 TWSPSKARLHOGSNAMRPQVNNREMLQVDFQTKMAYGVTTQYKSLISMTYKEFL 2280
QY 2281 ISSSQDGHQWTLFQNGKAVFQGNQDSETPVYNSLDPPLTRYLHPOSWHQTALRM 2340
DB 2281 ISSSQDGHQWTLFQNGKAVFQGNQDSETPVYNSLDPPLTRYLHPOSWHQTALRM 2340
QY 2341 EYVGEADLY 2351
DB 2341 EYVGEADLY 2351

RESULT 3
US-08-121-202-2
Sequence 2, Application US/08121202
Patent No. 5563045
GENERAL INFORMATION:
APPLICANT: Pitman, Debra
APPLICANT: Rehmtulla, Alnawaz
APPLICANT: Wozney, John M.
APPLICANT: Kaufman, Randal J.
TITLE OF INVENTION: CHIMERIC PROCOAGULANT PROTEINS
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: Legal Affairs, Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: MA
COUNTRY: USA
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/121,202
FILING DATE: 14-SEP-1993
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:
NAME: Melinert, M. C.
REGISTRATION NUMBER: 31,544
REFERENCE/DOCKET NUMBER: GI 5195A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 876-1210 X8574
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2351 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-121-202-2
Query Match 99.98% Score 12405; DB 1; Length 2351;
Best Local Similarity 99.98% Pred. No. 0;
Matches 2349; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 MOETLSTCEFLCLLNFCSATRRYYIGAVELSMQSDGLGELVDARPPRPVKSFPFN 60
DB 1 MOETLSTCEFLCLLNFCSATRRYYIGAVELSMQSDGLGELVDARPPRPVKSFPFN 60
QY 61 TSYVYKKTLEFETDHLFNIAKPRPMGLGPTIOAEVYDTVYTLKNNASHPVSLHAY 120
DB 61 TSYVYKKTLEFETDHLFNIAKPRPMGLGPTIOAEVYDTVYTLKNNASHPVSLHAY 120
QY 121 GVSVMKASGAEYDQISQREKEDKVPQGSHTYVWYKBNPMASDPLCTYSLSH 180
DB 121 GVSVMKASGAEYDQISQREKEDKVPQGSHTYVWYKBNPMASDPLCTYSLSH 180
QY 181 VDVVKNLNGILGALLVCREGSLAKERTQTLHKEFILFEVDEGKSMHSETKNSLMQDD 240
DB 181 VDVVKNLNGILGALLVCREGSLAKERTQTLHKEFILFEVDEGKSMHSETKNSLMQDD 240
QY 241 AASARAPKMHVNGYVRSPLGLGCHKRSYVNHYGSTPEVHSTLEGHTEFLVRNH 300
DB 241 AASARAPKMHVNGYVRSPLGLGCHKRSYVNHYGSTPEVHSTLEGHTEFLVRNH 300
QY 301 ROASLEISPTLEAQTLLMDIGQFLFLFCHISSHQDGEAAYKVDSCPEBPOLRMKNE 360
DB 301 ROASLEISPTLEAQTLLMDIGQFLFLFCHISSHQDGEAAYKVDSCPEBPOLRMKNE 360
QY 361 EADYDDDLTDESMQVRFDDONSPSFIQISYAKKHPTWVHTYAAEEDMOPAYPLTA 420
DB 361 EADYDDDLTDESMQVRFDDONSPSFIQISYAKKHPTWVHTYAAEEDMOPAYPLTA 420
QY 421 PDDRSYKSOYLNNGPQIRGKRYKRYEMAYTDEFTKRAIOHESGILGPLVGEVDTL 480
DB 421 PDDRSYKSOYLNNGPQIRGKRYKRYEMAYTDEFTKRAIOHESGILGPLVGEVDTL 480
QY 481 LITFKNQSHPYNIYPHGTIVRPLYSRRLPKGVKHLKDFPLPEIFRYKYWTVYEDGP 540
DB 481 LITFKNQSHPYNIYPHGTIVRPLYSRRLPKGVKHLKDFPLPEIFRYKYWTVYEDGP 540
QY 541 TKSDPRLCTRYYSSEFVNNERDLASGLIGPLLIYCKESVDQRNQIMSDKRNVLLEVFDE 600
DB 541 TKSDPRLCTRYYSSEFVNNERDLASGLIGPLLIYCKESVDQRNQIMSDKRNVLLEVFDE 600
QY 601 NRSWYLTENIORFLPPAGVQLEDEPEQASINHSINGVYFSDLSVCLLEVAVYVITLS 660
DB 601 NRSWYLTENIORFLPPAGVQLEDEPEQASINHSINGVYFSDLSVCLLEVAVYVITLS 660
QY 661 IGAQDTLVSFSGYTFKKHMYEDTLTLPPFSGEYVFMENPDLMTLIGCHNDFERNRG 720
DB 661 IGAQDTLVSFSGYTFKKHMYEDTLTLPPFSGEYVFMENPDLMTLIGCHNDFERNRG 720
QY 721 MTALLKYSQDKNTGNYEDSEYEDISAYLSKNNALPEPFSQNSRHSSTROKFNATTI 780
DB 721 MTALLKYSQDKNTGNYEDSEYEDISAYLSKNNALPEPFSQNSRHSSTROKFNATTI 780
QY 781 PENDIEKTDWEFAHRTPMKIQNVSSDDLMLLRQSTPFGHLSLSDLEAKYETFSDDPS 840
DB 781 PENDIEKTDWEFAHRTPMKIQNVSSDDLMLLRQSTPFGHLSLSDLEAKYETFSDDPS 840

```

Db      781 PENDIEKIDPWFHRTPMFKIQNVSSDDLMLRKOSP7PHGLSLSDLOEAKYETFEFDDPS 840
QY      841 PGADISNNSLSBMTFRPOLHSGDMVFPPESGLQRLNENKGT7TAATBLKTIKIDFVYST 900
Db      841 PGADISNNSLSBMTFRPOLHSGDMVFPPESGLQRLNENKGT7TAATBLKTIKIDFVYST 900
QY      901 SNNLSTIPSDMLAGTNTSSLOPPSPMPVHDSQDITTLTGKAKSSPLTSSGGPGLSLSE 960
Db      901 SNNLSTIPSDMLAGTNTSSLOPPSPMPVHDSQDITTLTGKAKSSPLTSSGGPGLSLSE 960
QY      961 NNDKSLLEGLANSOESSMGKNVSTESGRLFGKRAHGPALITKDNALFKVSIISLKTN 1020
Db      961 NNDKSLLEGLANSOESSMGKNVSTESGRLFGKRAHGPALITKDNALFKVSIISLKTN 1020
QY      1021 KTSNSATNRKTHIDGPSLLENSPWOMLIESDPEFEKATPLIHBMMDMKNATLAL 1080
Db      1021 KTSNSATNRKTHIDGPSLLENSPWOMLIESDPEFEKATPLIHBMMDMKNATLAL 1080
QY      1081 NMSNKTSSKNMKNVQOKKEGPTPPDAONPDMSFPKMLFLPESAKMTORTGKNSLNG 1140
Db      1081 NMSNKTSSKNMKNVQOKKEGPTPPDAONPDMSFPKMLFLPESAKMTORTGKNSLNG 1140
QY      1141 QGSPKQVLSLGPERSVEGQNFISEKNKVVVGKEFTKDVGLKEKVPSSNMLFLTNLDN 1200
Db      1141 QGSPKQVLSLGPERSVEGQNFISEKNKVVVGKEFTKDVGLKEKVPSSNMLFLTNLDN 1200
QY      1201 LHEHNTHNOEKKIOEIEIKKETLQIENVVLPQIHVTVGTAKNFMKMLFLSTRONVGSYD 1260
Db      1201 LHEHNTHNOEKKIOEIEIKKETLQIENVVLPQIHVTVGTAKNFMKMLFLSTRONVGSYD 1260
QY      1261 GAVAPYLOPRLINSTRKTKHAHNSKKEGEBNLEGLAQTOYIEKACTTRISPT 1320
Db      1261 GAVAPYLOPRLINSTRKTKHAHNSKKEGEBNLEGLAQTOYIEKACTTRISPT 1320
QY      1321 SQONFYQSRKALKQFRLPLEETELKRIIVDTSTOMSKMML7PSTLQIDYNEKE 1380
Db      1321 SQONFYQSRKALKQFRLPLEETELKRIIVDTSTOMSKMML7PSTLQIDYNEKE 1380
QY      1381 KGATOSPLSDCLTFRSHSIPQANSRPLIAKVSFSPISIPYLFVLFQONSSHLSAAY 1440
Db      1381 KGATOSPLSDCLTFRSHSIPQANSRPLIAKVSFSPISIPYLFVLFQONSSHLSAAY 1440
QY      1441 RKDQGVQESSHFLQOAKNNLSLAILTLMTGDQREVSIGTSAVNSVYKKEVNTVLP 1500
Db      1441 RKDQGVQESSHFLQOAKNNLSLAILTLMTGDQREVSIGTSAVNSVYKKEVNTVLP 1500
QY      1501 KPDLPTSGKVELLPRVHIYOKDLFPTETSNQSGHLDLVEGSLQTEGAIKMNANRP 1560
Db      1501 KPDLPTSGKVELLPRVHIYOKDLFPTETSNQSGHLDLVEGSLQTEGAIKMNANRP 1560
QY      1561 GKVPFLRVATESSAKTPSKLIDPLAMNHNCGQIPEEKKSQESPEKTAFAKKKIDTILSL 1620
Db      1561 GKVPFLRVATESSAKTPSKLIDPLAMNHNCGQIPEEKKSQESPEKTAFAKKKIDTILSL 1620
QY      1621 NACESNHAITAINEQONKPEIETVMAQGTENLCSQNPVLAKRHOREITRTTLQSDQEE 1680
Db      1621 NACESNHAITAINEQONKPEIETVMAQGTENLCSQNPVLAKRHOREITRTTLQSDQEE 1680
QY      1681 IDYDDTISYEMKKEFDIYDEENQSPRSFOKTRHFIYAAERLMDYGMSSSPHYLRNR 1740
Db      1681 IDYDDTISYEMKKEFDIYDEENQSPRSFOKTRHFIYAAERLMDYGMSSSPHYLRNR 1740
QY      1741 AQSGVPOFKKVVPOEFDGFTOPLYRGLNHLGLLGPYIRAEVEDNIMVTFRQASR 1800
Db      1741 AQSGVPOFKKVVPOEFDGFTOPLYRGLNHLGLLGPYIRAEVEDNIMVTFRQASR 1800
QY      1801 PYSFYSLSLISYEDDROGAEBRKNFVKNETKTYFWKVOHMAPTKEDFCKAMAFYSYV 1860
Db      1801 PYSFYSLSLISYEDDROGAEBRKNFVKNETKTYFWKVOHMAPTKEDFCKAMAFYSYV 1860
QY      1861 DLEKDYHSLGLIPLVCHNTLNPAGHROVTVQEFALFTJEDDKSMYETLENMENCRA 1920
Db      1861 DLEKDYHSLGLIPLVCHNTLNPAGHROVTVQEFALFTJEDDKSMYETLENMENCRA 1920

```

```

QY      1921 PCNIOMEDPFEKENREHANGIYIMDTLPGLVMAODORIRWITLMSGNSNIHSHIFSGH 1980
Db      1921 PCNIOMEDPFEKENREHANGIYIMDTLPGLVMAODORIRWITLMSGNSNIHSHIFSGH 1980
QY      1981 VFYVRKKEEKALNLYPGVFETVENLPSKAGIARVECLIGEHLHAGMSTLFLVYSNKC 2040
Db      1981 VFYVRKKEEKALNLYPGVFETVENLPSKAGIARVECLIGEHLHAGMSTLFLVYSNKC 2040
QY      2041 QTPLGMAIGHIDPOITASGQYGOAMAKRLARHYSGGSINANSTKEPSPWIKVDLAPMII 2100
Db      2041 QTPLGMAIGHIDPOITASGQYGOAMAKRLARHYSGGSINANSTKEPSPWIKVDLAPMII 2100
QY      2101 HGIRTOGAKOFSSLYXIOFTIMYSLUGKQWYRGNSTGTLMPFQNDOSGIRKININ 2160
Db      2101 HGIRTOGAKOFSSLYXIOFTIMYSLUGKQWYRGNSTGTLMPFQNDOSGIRKININ 2160
QY      2161 PPIIARTIRLHPHYSISITLMBELMCCDLSNMPLGESKASISDAQITASSYFTNMFA 2220
Db      2161 PPIIARTIRLHPHYSISITLMBELMCCDLSNMPLGESKASISDAQITASSYFTNMFA 2220
QY      2221 TWSPSKARLHLQGRSANRPOVNNPKEMLOVDFOKTMKVTGVTGQVKSLLTSMYKREL 2280
Db      2221 TWSPSKARLHLQGRSANRPOVNNPKEMLOVDFOKTMKVTGVTGQVKSLLTSMYKREL 2280
QY      2281 ISSQDGHQWTLFPQNGKRYKFOGNDSPFPVNSLDPPLTRRYLRHPOSWYHOZALRM 2340
Db      2281 ISSQDGHQWTLFPQNGKRYKFOGNDSPFPVNSLDPPLTRRYLRHPOSWYHOZALRM 2340
QY      2341 EYLGCEHODLY 2351
Db      2341 EYLGCEHODLY 2351

RESULT 4
5422260-1
; Patent No. 5422260
; APPLICANT: KAUFMAN, RANDAL J.; PITTMAN, DEBRA D.; TOOLE, JOHN J.
; TITLE OF INVENTION: HUMAN FACTOR VIII: C MUTAINS
; NUMBER OF SEQUENCES: 15
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/883, 936
; FILING DATE: 15-MAY-1992
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: 279,485
; FILING DATE: 02-DEC-1988; 09-DEC-1986
; APPLICATION NUMBER: 939,658
; FILING DATE: 09-DEC-1986
; APPLICATION NUMBER: 932,767
; FILING DATE: 18-NOV-1986
; APPLICATION NUMBER: 868,410
; SEQ ID NO:1:
; LENGTH: 2351
5422260-1

Query Match      99.8%; Score 12396; Db 6; Length 2351;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2348; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY      1 MQILSTCFCLLRCFSATRRYTLGAVELSDWYQSDGLGELPYDARFPPRPKSEPPN 60
Db      1 MQILSTCFCLLRCFSATRRYTLGAVELSDWYQSDGLGELPYDARFPPRPKSEPPN 60
QY      61 TSVYKKTLEFVEETHLRFNIAKPRPPMGLLGPTIOAEVYDVVYTLKNNASHPVSLAAV 120
Db      61 TSVYKKTLEFVEETHLRFNIAKPRPPMGLLGPTIOAEVYDVVYTLKNNASHPVSLAAV 120
QY      121 GVSYWKASBGAEDDOOTGORKEDDKVPCGSHTYWQVTKENGPMA SDPLCTVYYSLSH 180
Db      121 GVSYWKASBGAEDDOOTGORKEDDKVPCGSHTYWQVTKENGPMA SDPLCTVYYSLSH 180
QY      181 VDLVKDLSGLIGALLVCRBSLAKERTOTLAKFILLEAVDEGKSHSESTKNSLMQDRD 240

```

181 VDLKDNLSGLIGALLVCRBSGLAKETQTLHFIILFAVPDEGSMSEKKNLSMDRD 240
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000

1261 GAYAPVLODFRSLNDSTNRKHTAHFSKGBEENLGLNOTKOYEVKACTRISPT 1320
1321
1322
1323
1324
1325
1326
1327
1328
1329
1330
1331
1332
1333
1334
1335
1336
1337
1338
1339
1340
1341
1342
1343
1344
1345
1346
1347
1348
1349
1350
1351
1352
1353
1354
1355
1356
1357
1358
1359
1360
1361
1362
1363
1364
1365
1366
1367
1368
1369
1370
1371
1372
1373
1374
1375
1376
1377
1378
1379
1380
1381
1382
1383
1384
1385
1386
1387
1388
1389
1390
1391
1392
1393
1394
1395
1396
1397
1398
1399
1400
1401
1402
1403
1404
1405
1406
1407
1408
1409
1410
1411
1412
1413
1414
1415
1416
1417
1418
1419
1420
1421
1422
1423
1424
1425
1426
1427
1428
1429
1430
1431
1432
1433
1434
1435
1436
1437
1438
1439
1440
1441
1442
1443
1444
1445
1446
1447
1448
1449
1450
1451
1452
1453
1454
1455
1456
1457
1458
1459
1460
1461
1462
1463
1464
1465
1466
1467
1468
1469
1470
1471
1472
1473
1474
1475
1476
1477
1478
1479
1480
1481
1482
1483
1484
1485
1486
1487
1488
1489
1490
1491
1492
1493
1494
1495
1496
1497
1498
1499
1500
1501
1502
1503
1504
1505
1506
1507
1508
1509
1510
1511
1512
1513
1514
1515
1516
1517
1518
1519
1520
1521
1522
1523
1524
1525
1526
1527
1528
1529
1530
1531
1532
1533
1534
1535
1536
1537
1538
1539
1540
1541
1542
1543
1544
1545
1546
1547
1548
1549
1550
1551
1552
1553
1554
1555
1556
1557
1558
1559
1560
1561
1562
1563
1564
1565
1566
1567
1568
1569
1570
1571
1572
1573
1574
1575
1576
1577
1578
1579
1580
1581
1582
1583
1584
1585
1586
1587
1588
1589
1590
1591
1592
1593
1594
1595
1596
1597
1598
1599
1600
1601
1602
1603
1604
1605
1606
1607
1608
1609
1610
1611
1612
1613
1614
1615
1616
1617
1618
1619
1620
1621
1622
1623
1624
1625
1626
1627
1628
1629
1630
1631
1632
1633
1634
1635
1636
1637
1638
1639
1640
1641
1642
1643
1644
1645
1646
1647
1648
1649
1650
1651
1652
1653
1654
1655
1656
1657
1658
1659
1660
1661
1662
1663
1664
1665
1666
1667
1668
1669
1670
1671
1672
1673
1674
1675
1676
1677
1678
1679
1680
1681
1682
1683
1684
1685
1686
1687
1688
1689
1690
1691
1692
1693
1694
1695
1696
1697
1698
1699
1700
1701
1702
1703
1704
1705
1706
1707
1708
1709
1710
1711
1712
1713
1714
1715
1716
1717
1718
1719
1720
1721
1722
1723
1724
1725
1726
1727
1728
1729
1730
1731
1732
1733
1734
1735
1736
1737
1738
1739
1740
1741
1742
1743
1744
1745
1746
1747
1748
1749
1750
1751
1752
1753
1754
1755
1756
1757
1758
1759
1760
1761
1762
1763
1764
1765
1766
1767
1768
1769
1770
1771
1772
1773
1774
1775
1776
1777
1778
1779
1780
1781
1782
1783
1784
1785
1786
1787
1788
1789
1790
1791
1792
1793
1794
1795
1796
1797
1798
1799
1800
1801
1802
1803
1804
1805
1806
1807
1808
1809
1810
1811
1812
1813
1814
1815
1816
1817
1818
1819
1820
1821
1822
1823
1824
1825
1826
1827
1828
1829
1830
1831
1832
1833
1834
1835
1836
1837
1838
1839
1840
1841
1842
1843
1844
1845
1846
1847
1848
1849
1850
1851
1852
1853
1854
1855
1856
1857
1858
1859
1860
1861
1862
1863
1864
1865
1866
1867
1868
1869
1870
1871
1872
1873
1874
1875
1876
1877
1878
1879
1880
1881
1882
1883
1884
1885
1886
1887
1888
1889
1890
1891
1892
1893
1894
1895
1896
1897
1898
1899
1900
1901
1902
1903
1904
1905
1906
1907
1908
1909
1910
1911
1912
1913
1914
1915
1916
1917
1918
1919
1920
1921
1922
1923
1924
1925
1926
1927
1928
1929
1930
1931
1932
1933
1934
1935
1936
1937
1938
1939
1940
1941
1942
1943
1944
1945
1946
1947
1948
1949
1950
1951
1952
1953
1954
1955
1956
1957
1958
1959
1960
1961
1962
1963
1964
1965
1966
1967
1968
1969
1970
1971
1972
1973
1974
1975
1976
1977
1978
1979
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000

RESULT 5
US-07-864-004B-4
; Sequence 4, Application US/07864004B
; Patent No. 5364771
; GENERAL INFORMATION:
; APPLICANT: Iolair, John S.
; APPLICANT: Runge, Marshall S.
; TITLE OF INVENTION: Hybrid Human/Porcine Factor VIII
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kilpatrick & Cody
; STREET: 1100 Peachtree Street
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: US
; ZIP: 30309
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/864,004B
; FILING DATE: 07 April 1992
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Padst, Patrea L.
; REGISTRATION NUMBER: 31,284
; REFERENCE/DOCKET NUMBER: EMTU106
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404-815-6508
; TELEFAX: 404-815-6555
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2332 amino acids
; type: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; AMTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORGANISM: Homo sapien
; TISSUE TYPE: Liver CDNA sequence
; US-07-864-004B-4

Query Match 99.18; Score 12301; DB 1; Length 2332;
Best Local Similarity 99.98; Pred. No. 0;
Matches 2330; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```
QY 20 ATRRRYLGAVELSDWYQSDLDGELPYDARPPPRVKSFPFNTSVYKKTLFEFTDLN 79
DB 1 ATRRRYLGAVELSDWYQSDLDGELPYDARPPPRVKSFPFNTSVYKKTLFEFTDLN 60
QY 80 IAKPRPMWGLGPTIOAEVYDVVVTILKNMASHVSLHAAGVSYKAAASGAEYDQTSQ 139
DB 61 IAKPRPMWGLGPTIOAEVYDVVVTILKNMASHVSLHAAGVSYKAAASGAEYDQTSQ 120
QY 140 REKEDKVPFGSGHTYVQVLAKEGPMASDPLCTLYSYLSHVLDVKDLNSGLIGALLVR 199
DB 121 REKEDKVPFGSGHTYVQVLAKEGPMASDPLCTLYSYLSHVLDVKDLNSGLIGALLVR 180
QY 200 EGSIAKETOTLHKFTLLFAVPDEGKSMHSETKNSLMODRDAASARAWPMHATVNYNR 259
DB 181 EGSIAKETOTLHKFTLLFAVPDEGKSMHSETKNSLMODRDAASARAWPMHATVNYNR 240
QY 260 SLPLGLGCHRRSVYWHVAGGTPEVHSLFEGHTFLVRNHRQASLEISPTFLTAQTL 319
DB 241 SLPLGLGCHRRSVYWHVAGGTPEVHSLFEGHTFLVRNHRQASLEISPTFLTAQTL 300
```

```
QY 320 MDLGGFLLECHLSHNDGMEAYVYKVDSCPEERPOLRMKNEADYDDDLTDEMDVNF 379
DB 301 MDLGGFLLECHLSHNDGMEAYVYKVDSCPEERPOLRMKNEADYDDDLTDEMDVNF 360
QY 380 DDNSPSTQIRSYAKKHPTWVHYIAAEEDMDYAPLVLAADDRSYKSOYLNNKFORIG 439
DB 361 DDNSPSTQIRSYAKKHPTWVHYIAAEEDMDYAPLVLAADDRSYKSOYLNNKFORIG 420
QY 440 RYKKYRFMATYDETFRTRAIQHEGSIIGPLLYGEVDDTLIIIFKNQASRPNTYPHCI 499
DB 421 RYKKYRFMATYDETFRTRAIQHEGSIIGPLLYGEVDDTLIIIFKNQASRPNTYPHCI 480
QY 500 TVRPLYSRLPKGVKHLKDFPLPEEIFKYVTVYEDGPKSDPRCLTRYSSFVNE 559
DB 481 TVRPLYSRLPKGVKHLKDFPLPEEIFKYVTVYEDGPKSDPRCLTRYSSFVNE 540
QY 560 RDLASGLIGPLLYCKESVDQROGNOJMSKRNVILFSVDENRSMYLTENTIOEFLPNAG 619
DB 541 RDLASGLIGPLLYCKESVDQROGNOJMSKRNVILFSVDENRSMYLTENTIOEFLPNAG 600
QY 620 VLEDPPEQASNIHNSINGYVDSIQLSVCLHEVAYWYIISGAQTDPLSVFSGYTFKH 679
DB 601 VLEDPPEQASNIHNSINGYVDSIQLSVCLHEVAYWYIISGAQTDPLSVFSGYTFKH 660
QY 680 KMYEDTLTLFPFSGEYVFMSEMPGLMTIGCHNSDEFNRGNTALLKVSXCDKNTGDYE 739
DB 661 KMYEDTLTLFPFSGEYVFMSEMPGLMTIGCHNSDEFNRGNTALLKVSXCDKNTGDYE 720
QY 740 DSEYDISAYLLSKNAIEPRFSQNSRHPSTROKOFNATYIIPENDIEKDDPWFARHPMP 799
DB 721 DSEYDISAYLLSKNAIEPRFSQNSRHPSTROKOFNATYIIPENDIEKDDPWFARHPMP 780
QY 800 KIONVSSDLMLRQSPTRPHGLSLSDLOEAEYEFSSDPPGADISNNSLSMTHPRQ 859
DB 781 KIONVSSDLMLRQSPTRPHGLSLSDLOEAEYEFSSDPPGADISNNSLSMTHPRQ 840
QY 860 LHSQDMFTPEESGLQRLNEKLGTAATLKKLDFKYSTNNLSTIPSDNLAAGTDN 919
DB 841 LHSQDMFTPEESGLQRLNEKLGTAATLKKLDFKYSTNNLSTIPSDNLAAGTDN 900
QY 920 TSSLGPPSPMPVHYDSQDITLFGKSSPLTESGCLSLSEENDSKLLESGMLNQSSEM 979
DB 901 TSSLGPPSPMPVHYDSQDITLFGKSSPLTESGCLSLSEENDSKLLESGMLNQSSEM 960
QY 980 GKNVSTFSGRLFGKRAHGPALITDQNALPRVSLSLKTRKSNNSATNRKTHIDGSL 1039
DB 961 GKNVSTFSGRLFGKRAHGPALITDQNALPRVSLSLKTRKSNNSATNRKTHIDGSL 1020
QY 1040 LIENSPYMONLIESDTEFEKVTPLIHDRLMDKNATATRLNHSNKTTSKRMEMVOOK 1099
DB 1021 LIENSPYMONLIESDTEFEKVTPLIHDRLMDKNATATRLNHSNKTTSKRMEMVOOK 1080
QY 1100 KEGIPPPDAONPMSFPFKMLFLPESARWIOGRHKNSLNSGQSPKQVSLGPEKSYEG 1159
DB 1081 KEGIPPPDAONPMSFPFKMLFLPESARWIOGRHKNSLNSGQSPKQVSLGPEKSYEG 1140
QY 1160 QNFLEKKNVYVKGGEFTYDVGILEKEMVPPSSRNLFNLNLTENNTHNQEKIOEBIEK 1219
DB 1141 QNFLEKKNVYVKGGEFTYDVGILEKEMVPPSSRNLFNLNLTENNTHNQEKIOEBIEK 1200
QY 1220 KETLIOENVVLPOIHTYTGKKNPMKNTFLSTRNVGSDVGAAYAVLDQFSLNDSTNR 1279
DB 1201 KETLIOENVVLPOIHTYTGKKNPMKNTFLSTRNVGSDVGAAYAVLDQFSLNDSTNR 1260
QY 1280 TKKHTAHPSKKGEBNLEGLAQOTQOIVEKYACTRTISPNTSOQNFVTRSRKRLAKQRL 1339
DB 1261 TKKHTAHPSKKGEBNLEGLAQOTQOIVEKYACTRTISPNTSOQNFVTRSRKRLAKQRL 1320
QY 1340 PLEETLEKRLIYDDSTQMSKMKHLPPSLTQIDVNEKEKAIQSPLSCLTRSHSI 1399
DB 1321 PLEETLEKRLIYDDSTQMSKMKHLPPSLTQIDVNEKEKAIQSPLSCLTRSHSI 1380
QY 1400 PQANRSLPLAKVSEFSPSIRPIYTLRVLEFDNSSHLPAASVRRKDSGVQESSHFLQAKR 1459
```

1381 PQANRSPPLIAKVSSEFISIRPIYLFVLFODSSHLPAASARKKDSGVSSSHFLOQAKK 1440
1460 NNLSTALITLMTDQEVSGISGTATNSVYKKKVENTVLPKPDLPKTSKVELLPKPHI 1519
1441 NNLSTALITLMTDQEVSGISGTATNSVYKKKVENTVLPKPDLPKTSKVELLPKPHI 1500
1520 YOKDLFPTETNSGSPGHLDVBSGLLOQTEGAIKWNENRPAKVPPLKVAETESSAKTPSK 1579
1501 YOKDLFPTETNSGSPGHLDVBSGLLOQTEGAIKWNENRPAKVPPLKVAETESSAKTPSK 1560
1580 LLDPLAMNHYGTOIPKEEWSGSEKSPETAKPKKDDTLSTINACESVHAIAAINEGONKP 1639
1561 LLDPLAMNHYGTOIPKEEWSGSEKSPETAKPKKDDTLSTINACESVHAIAAINEGONKP 1620
1640 ELEVTAQOGTERLCSQNPVLAKRQREITVTLQSDQEEIDYDTISYEMKKEDEFDY 1699
1621 ELEVTAQOGTERLCSQNPVLAKRQREITVTLQSDQEEIDYDTISYEMKKEDEFDY 1680
1700 DEDENQSPRSFQKTRHYFAAVERLMDYGMSSSPHYLRANQSGSVPOPKKVVPOEFTD 1759
1681 DEDENQSPRSFQKTRHYFAAVERLMDYGMSSSPHYLRANQSGSVPOPKKVVPOEFTD 1740
1760 GSFQPLRYRGLNHLGLLPYRAVEDNIMVTRNQAASRPSYSLSISTEEDQROGA 1819
1741 GSFQPLRYRGLNHLGLLPYRAVEDNIMVTRNQAASRPSYSLSISTEEDQROGA 1800
1820 EPRKNEVAPNETKTYFVKVOHNAPTKDEFDCKANAYPSDVLDEKDVHSLIGPLVCHT 1879
1801 EPRKNEVAPNETKTYFVKVOHNAPTKDEFDCKANAYPSDVLDEKDVHSLIGPLVCHT 1860
1880 NTLNPAHROVTOEPLFTFIDETKSMVTEENNERCRAPCIOMEDTPEKENTRPA 1939
1861 NTLNPAHROVTOEPLFTFIDETKSMVTEENNERCRAPCIOMEDTPEKENTRPA 1920
1940 INGYIMDTLPGLVAAQDRIKWTLLSMGSNENIHSIHFSGHVTVRKKEEYKMAALYNLP 1999
1921 INGYIMDTLPGLVAAQDRIKWTLLSMGSNENIHSIHFSGHVTVRKKEEYKMAALYNLP 1980
2000 GVFEIVEMLPKAGIHWRECLIGEHLHAGMSTLFTVYNNKQOTPLGMAASHIDPQITAS 2059
1981 GVFEIVEMLPKAGIHWRECLIGEHLHAGMSTLFTVYNNKQOTPLGMAASHIDPQITAS 2040
2060 GOYGMAPRLARLHSGSINAMSTKEPFSWIKVDLAPMIHGIKTQGANOKFSSLYISQ 2119
2041 GOYGMAPRLARLHSGSINAMSTKEPFSWIKVDLAPMIHGIKTQGANOKFSSLYISQ 2100
2120 FTIMYSLDGKKMOTYRGNSGTGLMAVFGVNDSSGIKHNIENPPIIARYIKLHPTHTSIS 2179
2101 FTIMYSLDGKKMOTYRGNSGTGLMAVFGVNDSSGIKHNIENPPIIARYIKLHPTHTSIS 2160
2180 TLRMELMGCDLNSGSMPLGMSKASISDAQITASSYFTNMFATWSPSKARLHLQGRSNMR 2239
2161 TLRMELMGCDLNSGSMPLGMSKASISDAQITASSYFTNMFATWSPSKARLHLQGRSNMR 2220
2240 POVNNPKEMLOYDFOKTMKATYGTQGVSKSLTSMYVEKFLISSODGHOVTLFPONGVY 2299
2221 POVNNPKEMLOYDFOKTMKATYGTQGVSKSLTSMYVEKFLISSODGHOVTLFPONGVY 2280
2300 KYFQGNDSFTPPVNSLDPPLRLRYLRHPOSWHQIALRMEVIGCAADLY 2331
2281 KYFQGNDSFTPPVNSLDPPLRLRYLRHPOSWHQIALRMEVIGCAADLY 2332

RESULT 6

US-08-251-937A-4

; Sequence 4, Application US/08251937A

; Patent No. 5583209

; GENERAL INFORMATION:

; APPLICANT: Lollar, John S.

; APPLICANT: Runge, Marschall S.

; TITLE OF INVENTION: Hybrid Human/Porcine Factor VIII

; NUMBER OF SEQUENCES: 10

CORRESPONDENCE ADDRESS:
ADDRESSEE: Kilpatrick & Cody
STREET: 1100 Peachtree Street
CITY: Atlanta
STATE: Georgia
COUNTRY: US
ZIP: 30309
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/251,937A
FILING DATE: 31-MAY-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/864,004
FILING DATE: 07-APR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Pratt, John S.
REGISTRATION NUMBER: 29,476
REFERENCE/DOCKET NUMBER: EMD106D1V
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404-815-6367
TELEFAX: 404-815-6555
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 2332 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
ORIGINAL SOURCE:
ORGANISM: Homo sapien
TISSUE TYPE: Liver cdna sequence
US-08-251-937A-4

Query Match 99.1%; Score 12301; DB 1; Length 2332;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2330; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

20 ATRRYVIGAVELSMVDSMDLGELEPVDAKRPVPSFPFNTSVYKKTLEVEETDHLFN 79
1 ATRRYVIGAVELSMVDSMDLGELEPVDAKRPVPSFPFNTSVYKKTLEVEETVHLFN 60
80 IAKRPPMGLLAPTLQAEVYDVYVITLKNASHPVSLHAGVSYWKADEGAEVDDQTSQ 139
61 IAKRPPMGLLAPTLQAEVYDVYVITLKNASHPVSLHAGVSYWKADEGAEVDDQTSQ 120
140 REKEDDKVFFGSGHTYVQVLEKENGPMASDPLCTVSYLSHVDLYKDNLSGLIGALLVCR 199
121 REKEDDKVFFGSGHTYVQVLEKENGPMASDPLCTVSYLSHVDLYKDNLSGLIGALLVCR 180
200 EGSIAKEKQTLKFTLLFAVDEKSKHSETKSLMORBDASAARPKKHTYNGVNR 259
181 EGSIAKEKQTLKFTLLFAVDEKSKHSETKSLMORBDASAARPKKHTYNGVNR 240
260 SLPLIGCHRRSYVHWVIGMTPEVHSIFLEGHTLVNRHNOASLEISPTFTLTAQTL 319
241 SLPLIGCHRRSYVHWVIGMTPEVHSIFLEGHTLVNRHNOASLEISPTFTLTAQTL 300
320 MDLGOPLFCGHISSHOHDEMEATYVVDSCPEEPOLMKKNEAEYDDDLVDSMDVYRF 379
301 MDLGOPLFCGHISSHOHDEMEATYVVDSCPEEPOLMKKNEAEYDDDLVDSMDVYRF 360
380 DDNSPSFIOIRSVAKKPKTWVHYIAAEEEDMDYAPVLAAPDSSYSQYLNNGPORIG 439
361 DDNSPSFIOIRSVAKKPKTWVHYIAAEEEDMDYAPVLAAPDSSYSQYLNNGPORIG 420

QY 440 RKTKVREMAVYDETEKTRERAIQIHESGILGPELLYGBVGDTLLIIIFKNQASRPYNIYPHGI 499
DB 421 RKTKVREMAVYDETEKTRERAIQIHESGILGPELLYGBVGDTLLIIIFKNQASRPYNIYPHGI 480
QY 500 TDVRLPYSRRLPKGVKHLKDFPLPGEIFRYKWTYVEDGPKSDPRCLRRYSSPVNNE 559
DB 481 TDVRLPYSRRLPKGVKHLKDFPLPGEIFRYKWTYVEDGPKSDPRCLRRYSSPVNNE 540
QY 560 RDLASGLIGPLLICKESVDQGNQOIMSDKRNVLFSVDENRSMYLTENIOEFLPNPAG 619
DB 541 RDLASGLIGPLLICKESVDQGNQOIMSDKRNVLFSVDENRSMYLTENIOEFLPNPAG 600
QY 620 VOLEDEPQASNIMISINGYFDSIAQSVCLHEVAWYILSIGAOTDLSVFSGYTFKH 679
DB 601 VOLEDEPQASNIMISINGYFDSIAQSVCLHEVAWYILSIGAOTDLSVFSGYTFKH 660
QY 680 KMYEDTLTLPFSGEYFMSMENPGMLILGCHNSDFRNRGHTALLVSSCDKNTGDYFE 739
DB 661 KMYEDTLTLPFSGEYFMSMENPGMLILGCHNSDFRNRGHTALLVSSCDKNTGDYFE 720
QY 740 DSTDIEDISAYLLSKNNAIEPRSEFSONSHPESTROKOFNATTIPENDIEKTDPMFAHRTPM 799
DB 721 DSTDIEDISAYLLSKNNAIEPRSEFSONSHPESTROKOFNATTIPENDIEKTDPMFAHRTPM 780
QY 800 KIQNVSSDILLMLLNQSPTPHGLISLSDIOAKYETESDDPSGAIIDSNNLSIEMTHFRPQ 859
DB 781 KIQNVSSDILLMLLNQSPTPHGLISLSDIOAKYETESDDPSGAIIDSNNLSIEMTHFRPQ 840
QY 860 LHSNMDVFTPESGLOLNLNKLTGTTAETKLDYFVSTNNLTISTISDMLAGTGN 919
DB 841 LHSNMDVFTPESGLOLNLNKLTGTTAETKLDYFVSTNNLTISTISDMLAGTGN 900
QY 920 TSSIGPSPMPVHYDSQDLDTLLEGGKSSPLTESGGPLISEENNDSKLLSEGLMNSOESSW 979
DB 901 TSSIGPSPMPVHYDSQDLDTLLEGGKSSPLTESGGPLISEENNDSKLLSEGLMNSOESSW 960
QY 980 GKNVSSSTESGRLFKGRRAHGPALLTKDNALFKVSIISLTKNTKSNSATNRKTHIGPSL 1039
DB 961 GKNVSSSTESGRLFKGRRAHGPALLTKDNALFKVSIISLTKNTKSNSATNRKTHIGPSL 1020
QY 1040 LIENSPSVMONLTLEDTEFEKKVPLIHDRMLDMKNATLRLIMHSNNTSSKKNMEVQOK 1099
DB 1021 LIENSPSVMONLTLEDTEFEKKVPLIHDRMLDMKNATLRLIMHSNNTSSKKNMEVQOK 1080
QY 1100 KEGPIPPDAQNDMSFFEMLELPESARMIQRTHGKNSILNSGGQSPKQVLSLGEPSVGB 1159
DB 1081 KEGPIPPDAQNDMSFFEMLELPESARMIQRTHGKNSILNSGGQSPKQVLSLGEPSVGB 1140
QY 1160 ONFLSEKNVYVVGKEFTKQVGLKEMVFPSSRNLFTJNLNHLHENTHNOEKIOEIERK 1219
DB 1141 ONFLSEKNVYVVGKEFTKQVGLKEMVFPSSRNLFTJNLNHLHENTHNOEKIOEIERK 1200
QY 1220 KETLIOENNVLPQIHTVGTGNFKNLFILSTRQNVESGYDGAAPVLODFRSLANDSTR 1279
DB 1201 KETLIOENNVLPQIHTVGTGNFKNLFILSTRQNVESGYDGAAPVLODFRSLANDSTR 1260
QY 1280 TKKHTAHFSKKEEENLEGLGNQTKQIYEKACTYRISPTNSOQNFYORSKRALKOFR 1339
DB 1261 TKKHTAHFSKKEEENLEGLGNQTKQIYEKACTYRISPTNSOQNFYORSKRALKOFR 1320
QY 1340 PLEETELERKRIIVDSTOMSKNMKHLTPSTLOIDYNEKKGALITOSPISDCLTSHSI 1399
DB 1321 PLEETELERKRIIVDSTOMSKNMKHLTPSTLOIDYNEKKGALITOSPISDCLTSHSI 1380
QY 1400 POANRSPPLIAYVSSPISIRIYITRVLFODNSHLLPAASYRKKDSGVQESSHFLGAKK 1459
DB 1381 POANRSPPLIAYVSSPISIRIYITRVLFODNSHLLPAASYRKKDSGVQESSHFLGAKK 1440
QY 1460 NNLSTAILTLEMTGQREVGSIGTSATNSVTYKVENYVLPKPDLPKTSKVELLPKVI 1519
DB 1441 NNLSTAILTLEMTGQREVGSIGTSATNSVTYKVENYVLPKPDLPKTSKVELLPKVI 1500
QY 1520 YOKDLEPTEESNGSGHDLIVEGSLLOGTGGAIKMNEARPOKVEFLRVATESSATPSK 1579

DB 1501 YOKDLEPTEESNGSGHDLIVEGSLLOGTGGAIKMNEARPOKVEFLRVATESSATPSK 1560
QY 1580 LLDPLAMDNHGYGTOJPKREEMKSQSEKAPFAKKDITLISINACESHHAIAALINEONKP 1639
DB 1561 LLDPLAMDNHGYGTOJPKREEMKSQSEKAPFAKKDITLISINACESHHAIAALINEONKP 1620
QY 1640 EIEVYMAKQRTBERLCSQNPVYLKRHOEITRTTLOSDOEIDYDTISYEMKKEDEFDY 1699
DB 1621 EIEVYMAKQRTBERLCSQNPVYLKRHOEITRTTLOSDOEIDYDTISYEMKKEDEFDY 1680
QY 1700 DEDENQSPRSFOKTRRHYIAAVERLMDYGMSSSPHYLANRAGSGSPQPKVVFQEFND 1759
DB 1681 DEDENQSPRSFOKTRRHYIAAVERLMDYGMSSSPHYLANRAGSGSPQPKVVFQEFND 1740
QY 1760 GSFTOPLYRGELENEHGLGFTYIAEVEDNIMVTFRNOASRPSYTSLSISTYEDROGA 1819
DB 1741 GSFTOPLYRGELENEHGLGFTYIAEVEDNIMVTFRNOASRPSYTSLSISTYEDROGA 1800
QY 1820 EPRKNEVKNPEKTYFKRVQOHMAPTYKDEFDCKAMAYFSDVDLEKDVHSGILGPLLCHT 1879
DB 1801 EPRKNEVKNPEKTYFKRVQOHMAPTYKDEFDCKAMAYFSDVDLEKDVHSGILGPLLCHT 1860
QY 1880 NTLNPAHQVTVQEFALFTIPEDETSWTFENNERNCRAPCNIOHEDPTFKENRFP 1939
DB 1861 NTLNPAHQVTVQEFALFTIPEDETSWTFENNERNCRAPCNIOHEDPTFKENRFP 1920
QY 1940 INGYIMDPLGLVMAODORIRWYILSMGSEMENTHSIHPSGHYTVRKEEYKALYNLYP 1999
DB 1921 INGYIMDPLGLVMAODORIRWYILSMGSEMENTHSIHPSGHYTVRKEEYKALYNLYP 1980
QY 2000 GVEFVYBMLPSSAGIWRVECLIGELHAGNSTLELVYSNKCOTPLGMAASHIRDPQTAS 2059
DB 1981 GVEFVYBMLPSSAGIWRVECLIGELHAGNSTLELVYSNKCOTPLGMAASHIRDPQTAS 2040
QY 2060 GQYGOMAPKLARLHYSGSINAMSKPEPSIKVDLAPMIIHGILKTOGAROKSSYLYIO 2119
DB 2041 GQYGOMAPKLARLHYSGSINAMSKPEPSIKVDLAPMIIHGILKTOGAROKSSYLYIO 2100
QY 2120 FIIMYSIDGKKNQOTRGSTGTWVFFGNDSSGIGHNIFNPPIARIIRLHPHYSINS 2179
DB 2101 FIIMYSIDGKKNQOTRGSTGTWVFFGNDSSGIGHNIFNPPIARIIRLHPHYSINS 2160
QY 2180 TLKRELNGCDLNSCMPJGMSKASIDAOITASSYFTNNFATWSPSKARLHLQGRNMR 2239
DB 2161 TLKRELNGCDLNSCMPJGMSKASIDAOITASSYFTNNFATWSPSKARLHLQGRNMR 2220
QY 2240 POVANNPEMLQYDPQKTKKVTGVTGCVKSLTSMYVKEFLISSOGGHQWTLFPQNGV 2299
DB 2221 POVANNPEMLQYDPQKTKKVTGVTGCVKSLTSMYVKEFLISSOGGHQWTLFPQNGV 2280
QY 2300 KTFQGNDSFTPVVNSLDPPLITRYLRIHPQSVHOIALRMEVLGCEADOLY 2351
DB 2281 KTFQGNDSFTPVVNSLDPPLITRYLRIHPQSVHOIALRMEVLGCEADOLY 2332
RESULT 7
US-08-212-133A-2
; Sequence 2, Application US/08212133A
; Patent No. 5661060
; GENERAL INFORMATION:
; APPLICANT: Lollari, John S.
; APPLICANT: Kunge, Marschall S.
; TITLE OF INVENTION: Hybrid Human/Animal Factor VIII
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Kilpatrick & Cody
; STREET: 100 Peachtree Street
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: US
; ZIP: 30303
; COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/212,133A
FILING DATE: March 11, 1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/864,004
FILING DATE: 07-APR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Pabst, Patricia L.
REGISTRATION NUMBER: 31,284
REFERENCE/DOCKET NUMBER: EMU/76677
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404-572-6508
TELEFAX: 404-572-6555
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2332 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEICAL: YES
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
ORIGINAL SOURCE:
ORGANISM: Homo sapien
TISSUE TYPE: Liver CDNA sequence
US-08-212-133a-2

Query Match 99.18; Score 12301; DB 1; Length 2332;
Best Local Similarity 99.98; Pred. No. 0;
Matches 2330; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 20 ATRRYLLGAVELSDYQMSDGLGELPYDANFPYRVPKSEFNTSVYKTLFEVETDHLFN 79
DB 1 ATRRYLLGAVELSDYQMSDGLGELPYDANFPYRVPKSEFNTSVYKTLFEVETDHLFN 60
QY 80 IAKPRPMAGLIGPTIAEYVDVYVTLKNMASHVSLHANGSYMKASBEGAEYDDQSO 139
DB 61 IAKPRPMAGLIGPTIAEYVDVYVTLKNMASHVSLHANGSYMKASBEGAEYDDQSO 120
QY 140 REKEDKVPFGSHTYVWYVLKENGPMASDPLCTYVSHVDLVKDLNSGLGALLVCR 199
DB 121 REKEDKVPFGSHTYVWYVLKENGPMASDPLCTYVSHVDLVKDLNSGLGALLVCR 180
QY 200 EGSIAKEKTLQTLHFILLFAVDEGKSWHSETKNSLMODRDAASARAMPKHTVNGYVNR 259
DB 181 EGSIAKEKTLQTLHFILLFAVDEGKSWHSETKNSLMODRDAASARAMPKHTVNGYVNR 240
QY 260 SLPLGLIGHRKSVYWHYIGMGTPEVHSTLEEGHTELVKRNHQA3ETITTTTTCOTLE 319
DB 241 SLPLGLIGHRKSVYWHYIGMGTPEVHSTLEEGHTELVKRNHQA3ETITTTTTCOTLE 300
QY 320 MDLGGFLFCHISHQDGEAYVKKDSPEPTOTRKNNTEAEEDVDDLTJSEMDVVF 379
DB 301 MDLGGFLFCHISHQDGEAYVKKDSPEPTOTRKNNTEAEEDVDDLTJSEMDVVF 360
QY 380 DDNSPSFIOIRSAKKNHPTWVYIAAEEDMDYAPLYLAPDRSAYKSOYLNNPQRTG 439
DB 361 DDNSPSFIOIRSAKKNHPTWVYIAAEEDMDYAPLYLAPDRSAYKSOYLNNPQRTG 420
QY 440 RYKRYRPMATYDTEFTKRAIOHESGILGPILYEAGVDLILIFKNQASRYNITYPGI 499
DB 421 RYKRYRPMATYDTEFTKRAIOHESGILGPILYEAGVDLILIFKNQASRYNITYPGI 480
QY 500 TVRPLYSRRLPKGVKHLKDFLLPGLKIKWYVYVDEGPTKSDPRCLTRYSSFFVME 559
DB 481 TVRPLYSRRLPKGVKHLKDFLLPGLKIKWYVYVDEGPTKSDPRCLTRYSSFFVME 540

QY 560 RDLASGLIGPLLCYKESVDQNGQIMSDKRNVLFSVEDENRSTWLTENIQRELPNPAG 619
DB 541 RDLASGLIGPLLCYKESVDQNGQIMSDKRNVLFSVEDENRSTWLTENIQRELPNPAG 600
QY 620 VOLEDPERQASINMHSNGYVFDLSQVCLAEVAYVYLIISAOQDFLSVFSFGYFKH 679
DB 601 VOLEDPERQASINMHSNGYVFDLSQVCLAEVAYVYLIISAOQDFLSVFSFGYFKH 660
QY 680 KMYEDTLEFPESGEPVMSMNGIATLCHSDSPRNKGMATLAKVSCDKNKTGYE 739
DB 661 KMYEDTLEFPESGEPVMSMNGIATLCHSDSPRNKGMATLAKVSCDKNKTGYE 720
QY 740 DSYEDISAYLLSKNNALIEPRFSQNSRHSSTROKOFNATYIPENDIEKTPWFARHPMP 799
DB 721 DSYEDISAYLLSKNNALIEPRFSQNSRHSSTROKOFNATYIPENDIEKTPWFARHPMP 780
QY 800 KIONVSSDLMMLRQSPPHGLSLSDQEAKEVTEFDDSPCAIDSNNSLSEMTHPAPQ 859
DB 781 KIONVSSDLMMLRQSPPHGLSLSDQEAKEVTEFDDSPCAIDSNNSLSEMTHPAPQ 840
QY 860 LHSQDMVFTPEESGLQRLNEKLTGAATELKLDFVYSTSNMLISTPDSNLAAGTDN 919
DB 841 LHSQDMVFTPEESGLQRLNEKLTGAATELKLDFVYSTSNMLISTPDSNLAAGTDN 900
QY 920 TSLGPPSPMPHYDSQDITLFGKSSPLTESGSPSLSEENDSKLSEGLMNSQESSW 979
DB 901 TSLGPPSPMPHYDSQDITLFGKSSPLTESGSPSLSEENDSKLSEGLMNSQESSW 960
QY 980 GKNVSTSGRFLFKRAHAPALLTDNALKRVYSLSLTKNKTNNKATKRTHTDPSL 1039
DB 961 GKNVSTSGRFLFKRAHAPALLTDNALKRVYSLSLTKNKTNNKATKRTHTDPSL 1020
QY 1040 LTBNSBYQNLIESDTEFKKVTPLIHDRLMDNATLRLNHSKNTTSSKNEMVQOK 1099
DB 1021 LTBNSBYQNLIESDTEFKKVTPLIHDRLMDNATLRLNHSKNTTSSKNEMVQOK 1080
QY 1100 KEGPIPPDAQNDMSFFKMLFEPESARWIOETHKNSLNSGGQSPSKQVLSGPEKSYEG 1159
DB 1081 KEGPIPPDAQNDMSFFKMLFEPESARWIOETHKNSLNSGGQSPSKQVLSGPEKSYEG 1140
QY 1160 QNPLSEKKNVYVGGKPTKDYGLKEVNPSSNPLFTNLNLDHENTHOEKKIOEIEK 1219
DB 1141 QNPLSEKKNVYVGGKPTKDYGLKEVNPSSNPLFTNLNLDHENTHOEKKIOEIEK 1200
QY 1220 KETLQENNVLPQIHVYTGKKNMKNLFLSTRQNVGSGYGAAYAPVLQDFRSLNDSTNR 1279
DB 1201 KETLQENNVLPQIHVYTGKKNMKNLFLSTRQNVGSGYGAAYAPVLQDFRSLNDSTNR 1260
QY 1280 TKKHTAFHKKGEENLEGQNTQOIVEKYACTTRISPTSQOENVTORSKRALQPR 1339
DB 1261 TKKHTAFHKKGEENLEGQNTQOIVEKYACTTRISPTSQOENVTORSKRALQPR 1320
QY 1340 PLSETELEKRLIYVDSTQSKMKMLPSPILTQIDNEKEKATIQSLPSCLTRHSI 1399
DB 1321 PLSETELEKRLIYVDSTQSKMKMLPSPILTQIDNEKEKATIQSLPSCLTRHSI 1380
QY 1400 POANSPLPIAKVSSPSPIRPLYLTVFLJFODNSHLPASVRRKDSGVDESHFLOGAKK 1459
DB 1381 POANSPLPIAKVSSPSPIRPLYLTVFLJFODNSHLPASVRRKDSGVDESHFLOGAKK 1440
QY 1460 NNLSLATLLEMTDQGEVSGLSGTSNYSYVYKVENTVLPKRDLPKTSYGKELLPVYH 1519
DB 1441 NNLSLATLLEMTDQGEVSGLSGTSNYSYVYKVENTVLPKRDLPKTSYGKELLPVYH 1500
QY 1520 YOKDLPETISNGSPGHLDIYBGSLLQGTBGAIKWNPANPQVPLRYATBSSAKTPSK 1579
DB 1501 YOKDLPETISNGSPGHLDIYBGSLLQGTBGAIKWNPANPQVPLRYATBSSAKTPSK 1560
QY 1580 LADPLAMDNHGTQIPEEHWKSOEKSPEKFAKKKDDILSLNACSENHAIINAGQNK 1639
DB 1561 LADPLAMDNHGTQIPEEHWKSOEKSPEKFAKKKDDILSLNACSENHAIINAGQNK 1620
QY 1640 EIEVYVWAKGRTBRLCSQNPVYLKRNQREIIRTLQSDQDEIDYDITISVEKKEDDIT 1699


```

Db      1621 EIEVNAKOGKTERELCSNPPVLRKHORETTRTTLQSDKEIDYDXTSVMMKKEPDY 1680
QY      1700 DEDENOSPRSFOKTRHFIYAVERLMDYOMSSSPHYLNNRAOSGVSPOFKVYFOEFD 1759
Db      1681 DEDENOSPRSFOKTRHFIYAVERLMDYOMSSSPHYLNNRAOSGVSPOFKVYFOEFD 1740
QY      1760 GSFTOPLRGELNEHGLGPIYIAVEDNIMTFEPNOSRPSYSSLSYEEDNOGA 1819
Db      1741 GSFTOPLRGELNEHGLGPIYIAVEDNIMTFEPNOSRPSYSSLSYEEDNOGA 1800
QY      1820 EPRKNFVKNETKTYFMVQHBMAPTEDEFCANAYFSDVDLEKDVHSGILGFLVCHT 1879
Db      1801 EPRKNFVKNETKTYFMVQHBMAPTEDEFCANAYFSDVDLEKDVHSGILGFLVCHT 1860
QY      1880 NTLNPAHGRQVYQOEALFETTFDETSKMYFTENMERNCRAPCNIOEMDPFKENYFHA 1939
Db      1861 NTLNPAHGRQVYQOEALFETTFDETSKMYFTENMERNCRAPCNIOEMDPFKENYFHA 1920
QY      1940 INGIYNDTLPLGLVMAODORIRWYLLSGNSNENHSHFSGHVFYRKKEEKALYLYP 1999
Db      1921 INGIYNDTLPLGLVMAODORIRWYLLSGNSNENHSHFSGHVFYRKKEEKALYLYP 1980
QY      2000 GFETVEMLPKAGIWRVECLIGEHLAAGSTLELVYSNKCOTPLGMAHGHINDPQITAS 2059
Db      1981 GFETVEMLPKAGIWRVECLIGEHLAAGSTLELVYSNKCOTPLGMAHGHINDPQITAS 2040
QY      2060 GQYGMAPKLARLHSGSINAMSTKEPSMIRKDLLAPMIHIGITOGAROKFESLITQ 2119
Db      2041 GQYGMAPKLARLHSGSINAMSTKEPSMIRKDLLAPMIHIGITOGAROKFESLITQ 2100
QY      2120 FIIMYSLDGKKWQYNGSTGTAWFGVNDSSGIRHNIPNPITARIYRLHPHYSIRS 2179
Db      2101 FIIMYSLDGKKWQYNGSTGTAWFGVNDSSGIRHNIPNPITARIYRLHPHYSIRS 2160
QY      2180 TLRMELGCDLNSCNSPLGMEKASIDAOITASSYFTNMFATWSPSKARLHLOQRNMR 2239
Db      2161 TLRMELGCDLNSCNSPLGMEKASIDAOITASSYFTNMFATWSPSKARLHLOQRNMR 2220
QY      2240 POVNNPREMLOVDFOKTMVGTGOGKSLTSMYKFEILLSODDHOHTLFFONGKY 2299
Db      2221 POVNNPREMLOVDFOKTMVGTGOGKSLTSMYKFEILLSODDHOHTLFFONGKY 2280
QY      2300 KVFQGNDSFTPVVNSLDPLLTRYLRIHQSVWQIALRMEYLCEADLY 2351
Db      2281 KVFQGNDSFTPVVNSLDPLLTRYLRIHQSVWQIALRMEYLCEADLY 2332

RESULT 8
US-08-474-503-2
: Sequence 2. Application US/08474503
: Patent No. 5744446
: GENERAL INFORMATION:
: APPLICANT: Emory University
: TITLE OF INVENTION: Hybrid Human/Animal Factor VIII
: NUMBER OF SEQUENCES: 12
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Kilpatrick & Cody
: STREET: 1100 Peachtree Street, Suite 2800
: CITY: Atlanta
: STATE: Georgia
: COUNTRY: US
: ZIP: 30309
: COMPUTER READABLE FORM:
: MEDIUM TYPE: floppy disk
: OPERATING SYSTEM: IBM PC compatible
: SOFTWARE: Patent Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/474,503
: FILING DATE: 07-JUN-1995
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:

```

```

: NAME: Pratt, John S.
: REGISTRATION NUMBER: 29,476
: REFERENCE/DOCKET NUMBER: EM0106CIP(3)
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 404-815-6500
: TELEFAX: 404-815-6555
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2332 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: HYPOTHEICAL: YES
: ANTI-SENSE: NO
: FRAGMENT TYPE: N-terminal
: ORIGINAL SOURCE:
: ORGANISM: Homo sapien
: TISSUE TYPE: liver cDNA sequence
US-08-474-503-2

Query Match 99.1%; Score 12301; DB 1; Length 2332;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2330; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      20 ATRRYIYGAVELSMYQSDGLDELFPVDAFPKSPFNTSVYKTLFEFTDLFN 79
Db      1 ATRRYIYGAVELSMYQSDGLDELFPVDAFPKSPFNTSVYKTLFEFTDLFN 60
QY      80 IAKPRPPMGLIPTQAEVYTVYITLAKNASHPVSLAVGYSYWKASBEAEVDQTSQ 139
Db      61 IAKPRPPMGLIPTQAEVYTVYITLAKNASHPVSLAVGYSYWKASBEAEVDQTSQ 120
QY      140 REKEDKVPFGSGHTYVMQYLKENGPMASDPLCTYSYLSHVDLYKDLNSGLIGALLYCR 199
Db      121 REKEDKVPFGSGHTYVMQYLKENGPMASDPLCTYSYLSHVDLYKDLNSGLIGALLYCR 180
QY      200 EBSLAKETQTLKFFILFAVYDEKSWNSSTKNSLMQDPAASAPKKNHVTNGVYNR 259
Db      181 EBSLAKETQTLKFFILFAVYDEKSWNSSTKNSLMQDPAASAPKKNHVTNGVYNR 240
QY      260 SLPLGICGHRKSYVYVIGMGTTPVSHSLDEGHTPTTKRKAQSLSTSPITLTAQTL 319
Db      241 SLPLGICGHRKSYVYVIGMGTTPVSHSLDEGHTPTTKRKAQSLSTSPITLTAQTL 300
QY      320 MDLGFLLFCHISSHODGMEATVYVDSCPREPQLRRKKNNEAEVDYDGLTDEKDYVRF 379
Db      301 MDLGFLLFCHISSHODGMEATVYVDSCPREPQLRRKKNNEAEVDYDGLTDEKDYVRF 360
QY      380 DDDNSPSPFIOIRSVAKKHPKPTWVYHIAEEDMDYAPLYAPDRSRYSGOYLNNPQIRG 439
Db      361 DDDNSPSPFIOIRSVAKKHPKPTWVYHIAEEDMDYAPLYAPDRSRYSGOYLNNPQIRG 420
QY      440 KRYKVRPMAYTDETKTRDAIQHESGILGPLYGEVGDPLLIIFPNQASRPYNTYPHGI 499
Db      421 KRYKVRPMAYTDETKTRDAIQHESGILGPLYGEVGDPLLIIFPNQASRPYNTYPHGI 480
QY      500 TDVRLYSRRLPKGVKHLKDFPLPGELFKKATVYTEDPRTSDPRCLTRYSSFYVME 559
Db      481 TDVRLYSRRLPKGVKHLKDFPLPGELFKKATVYTEDPRTSDPRCLTRYSSFYVME 540
QY      560 RDLASGLGPLLITCYKESYVDQNGNDMDKRNVLFSVPEBNSWLTPTNORPLPNAG 619
Db      541 RDLASGLGPLLITCYKESYVDQNGNDMDKRNVLFSVPEBNSWLTPTNORPLPNAG 600
QY      620 VQLEDPERQASNIHNSINGVYEDSLOLSVCLAEVAYVYILISGAQDPLSVFSGYTFKH 679
Db      601 VQLEDPERQASNIHNSINGVYEDSLOLSVCLAEVAYVYILISGAQDPLSVFSGYTFKH 660
QY      680 KMYEDTLTLFPFSGEYVPMSENGMGLITGCHSDPRNKGATLAKVSSCCDKNTGYTE 729
Db      661 KMYEDTLTLFPFSGEYVPMSENGMGLITGCHSDPRNKGATLAKVSSCCDKNTGYTE 720

```

```
QY 740 DSYEDISAYILSKNNAIEPRSEFSQNSRHPSTROKOFNATTIPENDIEKTDPMFAHRTMP 799
DB 721 DSYEDISAYILSKNNAIEPRSEFSQNSRHPSTROKOFNATTIPENDIEKTDPMFAHRTMP 780
QY 800 KIQWSSDDLMLLRQSPTRPHGLSLSDIQEAKYETFDSDPSPADIDNSNLSSEMHFRQ 859
DB 781 KIQWSSDDLMLLRQSPTRPHGLSLSDIQEAKYETFDSDPSPADIDNSNLSSEMHFRQ 840
QY 860 LHSQDMWTFPEBSGLQIRLNEKLGTTAATELKLDFKVSSTSNLSTIPSDNLAAGTDN 919
DB 841 LHSQDMWTFPEBSGLQIRLNEKLGTTAATELKLDFKVSSTSNLSTIPSDNLAAGTDN 900
QY 920 TSSIGPSPMRYHDSODITLLEKSKSPITSGGGLSISEENDSKLSEGLMSQESSW 979
DB 901 TSSIGPSPMRYHDSODITLLEKSKSPITSGGGLSISEENDSKLSEGLMSQESSW 960
QY 980 GKNVSTESGRLFKGRABHAPALLTKDNALFKVSTSLKTKNTSNNATNRKTHIDGPSL 1039
DB 961 GKNVSTESGRLFKGRABHAPALLTKDNALFKVSTSLKTKNTSNNATNRKTHIDGPSL 1020
QY 1040 LSENSPYWONILSDPEFEKKTVPPLIHDMLMDKNAATRLNHSNKTTSSKMMEMVQOK 1099
DB 1021 LIENSPYWONILSDPEFEKKTVPPLIHDMLMDKNAATRLNHSNKTTSSKMMEMVQOK 1080
QY 1100 KEGPIPDANPDMSFFKMLLPESARWIOPTHGKNSLNSGQSPKQVSLGPEKSVBG 1159
DB 1081 KEGPIPDANPDMSFFKMLLPESARWIOPTHGKNSLNSGQSPKQVSLGPEKSVBG 1140
QY 1160 ONFLEKKNVYVGGEFTKDVGLKEMVFPSSRNLETLMDLJHENHNNHNEKKIOEIEK 1219
DB 1141 ONFLEKKNVYVGGEFTKDVGLKEMVFPSSRNLETLMDLJHENHNNHNEKKIOEIEK 1200
QY 1220 KETLIOENVVLPQIHTYTGKKNKMLFLSTRQVBSYGAAAPVLDQFRSLNJSIR 1260
DB 1201 KETLIOENVVLPQIHTYTGKKNKMLFLSTRQVBSYGAAAPVLDQFRSLNJSIR 1260
QY 1280 TKKHTAHFSKKGEEENLEGLNOTKOIVKRYACTRISPTNSQONFTQSRKALQFRL 1339
DB 1261 TKKHTAHFSKKGEEENLEGLNOTKOIVKRYACTRISPTNSQONFTQSRKALQFRL 1320
QY 1340 PLEETLEKRIITVDSTQWSKMKHLSTLTQIDYNEKKGATQSPISDCLTSHSI 1399
DB 1321 PLEETLEKRIITVDSTQWSKMKHLSTLTQIDYNEKKGATQSPISDCLTSHSI 1380
QY 1400 PQANRSPPLIAKVSFSPSIRIYLRVLFQDNSHLPAASTRKKDSGVQESSHFLQGAOK 1459
DB 1381 PQANRSPPLIAKVSFSPSIRIYLRVLFQDNSHLPAASTRKKDSGVQESSHFLQGAOK 1440
QY 1460 NNLSTALITLLEMTGQORVSGISGTSATNSVTYKRYENVLPKPLPTSGKVELLPKVI 1519
DB 1441 NNLSTALITLLEMTGQORVSGISGTSATNSVTYKRYENVLPKPLPTSGKVELLPKVI 1500
QY 1520 YOKDLFPETSNGBFGLDIVESSLQGTBGAIKMNEANRPQVPELRVATESAKTPK 1579
DB 1501 YOKDLFPETSNGBFGLDIVESSLQGTBGAIKMNEANRPQVPELRVATESAKTPK 1560
QY 1580 LLDPLAMDNHYGTOIPKREKMSQKSEKTAFFKKDTILSNACESHAAIAINEGONP 1639
DB 1561 LLDPLAMDNHYGTOIPKREKMSQKSEKTAFFKKDTILSNACESHAAIAINEGONP 1620
QY 1640 EIEVTWAKOGRTERLCQNPVYLRKHOREITRTTLOSDEELDDDTISYEMKKEFDIY 1699
DB 1621 EIEVTWAKOGRTERLCQNPVYLRKHOREITRTTLOSDEELDDDTISYEMKKEFDIY 1680
QY 1700 DEDENSPSPFOKTRHFIYAVERLMDYGMSSSPVLRNRAQSGVPOFKVYQOEDT 1759
DB 1681 DEDENSPSPFOKTRHFIYAVERLMDYGMSSSPVLRNRAQSGVPOFKVYQOEDT 1740
QY 1760 GSFTOPLRYGELNEHILGLPYIRAEVEDINWTFRQNASRPYSYSLSYEEDORQA 1819
DB 1741 GSFTOPLRYGELNEHILGLPYIRAEVEDINWTFRQNASRPYSYSLSYEEDORQA 1800
QY 1820 EPRKNFVKPNEFKTYFMWVQHMAFPTDEHDCAMAYFSDVLDKDVHSLIGLIVCHT 1879
```

```
DB 1801 EPRKNFVKPNEFKTYFMWVQHMAFPTDEHDCAMAYFSDVLDKDVHSLIGLIVCHT 1860
QY 1880 NTLNPAHQVQVCEFALEFTEDETSKMYETENNERNCRAPCNIOMEDEPTFKENYRFA 1939
DB 1861 NTLNPAHQVQVCEFALEFTEDETSKMYETENNERNCRAPCNIOMEDEPTFKENYRFA 1920
QY 1940 INGTMDTLPGLVMAQODRIRWYLLSGNSNENHSHHSQAVTFVRRKEEYKMAVNLVP 1999
DB 1921 INGTMDTLPGLVMAQODRIRWYLLSGNSNENHSHHSQAVTFVRRKEEYKMAVNLVP 1980
QY 2000 GVEFEYEMLPKRAQIMRVECLIGEBHLHAGNSTLFLVYSNKCOTPLGMASGHIDPOITAS 2059
DB 1981 GVEFEYEMLPKRAQIMRVECLIGEBHLHAGNSTLFLVYSNKCOTPLGMASGHIDPOITAS 2040
QY 2060 GQYGWAPPLAHLHSGSINMSTKEEFSIKYDLAPMTHIGKITQAGAROKTSYLISO 2119
DB 2041 GQYGWAPPLAHLHSGSINMSTKEEFSIKYDLAPMTHIGKITQAGAROKTSYLISO 2100
QY 2120 FIIMYSIDCKKQOTYRGNSTGTLWVFFGVNDSSGIKHNITNPPIARIYRLHPTHSIRS 2179
DB 2101 FIIMYSIDCKKQOTYRGNSTGTLWVFFGVNDSSGIKHNITNPPIARIYRLHPTHSIRS 2160
QY 2180 TIRHEIMGCDLNSGMPJGMSKAIISDAQITASSYFTNNFATWSPSKARHLQGRNMR 2239
DB 2161 TIRHEIMGCDLNSGMPJGMSKAIISDAQITASSYFTNNFATWSPSKARHLQGRNMR 2220
QY 2240 POWNPREMLQVDFQTKMKTGVTTQGVKSILTSMTYKEFLISSQDGHOWTFPQNGKY 2299
DB 2221 POWNPREMLQVDFQTKMKTGVTTQGVKSILTSMTYKEFLISSQDGHOWTFPQNGKY 2280
QY 2300 KVFQGNDSFTPVNSLDPLTIRLIRHPQSVWQIALRMEVLCEAODLY 2351
DB 2281 KVFQGNDSFTPVNSLDPLTIRLIRHPQSVWQIALRMEVLCEAODLY 2332
```

```
DB 2300 KVFQGNDSFTPVNSLDPLTIRLIRHPQSVWQIALRMEVLCEAODLY 2351
DB 2281 KVFQGNDSFTPVNSLDPLTIRLIRHPQSVWQIALRMEVLCEAODLY 2332
```

```
RESULT 9
US-08-670-707A-2
Sequence 2, Application US/08670707A
Patent No. 5859204
GENERAL INFORMATION:
Applicant: Iollar, John S.
TITLE OF INVENTION: Hybrid Human/Animal Factor VIII
NUMBER OF SEQUENCES: 40
CORRESPONDENCE ADDRESS:
ADDRESS: Greenlee, Winner and Sullivan, P.C.
STREET: 5370 Manhattan Circle Suite 201
CITY: Boulder
STATE: Colorado
COUNTRY: USA
ZIP: 80303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/670,707A
FILING DATE: 26-JUN-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: NO PCT/US94/13200
FILING DATE: 15-NOV-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/212,133
FILING DATE: 11-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/864,004
FILING DATE: 07-APR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Greenlee, Lorraine L.
REGISTRATION NUMBER: 27,894
REFERENCE/DOCKET NUMBER: 75-95F
```

```

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 303/499-8080
; TELEFAX: 303/499-8089
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2332 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; HYPOTHEICAL: YES
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; TISSUE TYPE: Liver
; US-08-670-707A-2

Query Match      99.1%; Score 12301; DB 2; Length 2332;
Best Local Similarity 99.9%; Fred. No. 0;
Matches 2330; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      20 ATRRRYLGAVELSDMYMOSDGLPVDARPPVPKSPFPNTSVYKKTLVEFTDHLN 79
      1 ATRRRYLGAVELSDMYMOSDGLPVDARPPVPKSPFPNTSVYKKTLVEFTDHLN 60
QY      80 IAKRPPMGLGPTIOAEYTDVYITLKNMASHPLAVGVSVYKASBGAEYDQTSQ 139
      61 IAKRPPMGLGPTIOAEYTDVYITLKNMASHPLAVGVSVYKASBGAEYDQTSQ 120
QY      140 REKEDKVPFGSGSHYVQVLTKEGPMASDPLCLTYSYSHVDVKNLSGLGALLVCR 199
      121 REKEDKVPFGSGSHYVQVLTKEGPMASDPLCLTYSYSHVDVKNLSGLGALLVCR 180
QY      200 EGSIAKEXTQTLHKETLLFAVFDGKSMHSETKNSLMODRDAASARAPKMHYNGYVR 259
      181 EGSIAKEXTQTLHKETLLFAVFDGKSMHSETKNSLMODRDAASARAPKMHYNGYVR 240
QY      260 SLPLGICHRKSYMHVIGMGTPPEVHSIFLEGHFLVKNHQALESPTPTLTQTL 319
      241 SLPLGICHRKSYMHVIGMGTPPEVHSIFLEGHFLVKNHQALESPTPTLTQTL 300
QY      320 MDLQOPLIFCHISSHOHGMETAYKVDSCBEPOLPKKNNEAEYDDDLTDSMDVYR 379
      301 MDLQOPLIFCHISSHOHGMETAYKVDSCBEPOLPKKNNEAEYDDDLTDSMDVYR 360
QY      380 DDNSPSFIQIRSVAKKPKTWVHYIAAEEEDMDYAPLVLAPODRSKSYQLANGPQRTG 439
      361 DDNSPSFIQIRSVAKKPKTWVHYIAAEEEDMDYAPLVLAPODRSKSYQLANGPQRTG 420
QY      440 RKKYKVFMAATDEFTKTRREALIOHESGILGPLLYXGEVGTLLITKNOASPRNYTPHGI 499
      421 RKKYKVFMAATDEFTKTRREALIOHESGILGPLLYXGEVGTLLITKNOASPRNYTPHGI 480
QY      500 TDVRLYSRRLPKGVKHLKDEPILPGLIFKYKWTVTVEDGPTKSDPRLCTRYYSFVNM 559
      481 TDVRLYSRRLPKGVKHLKDEPILPGLIFKYKWTVTVEDGPTKSDPRLCTRYYSFVNM 540
QY      560 RDLASGLIGPLICKEVSDOGRNOIMSDKRNVLFSVDEKRSYLTLENIOPLPAPAG 619
      541 RDLASGLIGPLICKEVSDOGRNOIMSDKRNVLFSVDEKRSYLTLENIOPLPAPAG 600
QY      620 VQLEDPEFOASNMHSINGYVDSLOLISVCLHEVAWYILISGAQTDLSFFSGYTFKH 679
      601 VQLEDPEFOASNMHSINGYVDSLOLISVCLHEVAWYILISGAQTDLSFFSGYTFKH 660
QY      680 KMYEDTLTLFPESGETVFMENPGMILGCHNSDFRNKGTALLKYSQCKWTGQYVE 739
      661 KMYEDTLTLFPESGETVFMENPGMILGCHNSDFRNKGTALLKYSQCKWTGQYVE 720
QY      740 DSYEDISAYILSKNNAIEPRSPSONSRHPTROKOPNATTIENDIETDWDVPHRPPMP 799
      721 DSYEDISAYILSKNNAIEPRSPSONSRHPTROKOPNATTIENDIETDWDVPHRPPMP 780

```

```

QY      800 KIQVSSDDLMLNRQSPPTPHGLSLDQAEKYEFPDSDPGAIDSNNSISEWTHFRPQ 859
      781 KIQVSSDDLMLNRQSPPTPHGLSLDQAEKYEFPDSDPGAIDSNNSISEWTHFRPQ 840
QY      860 LHSQDMVFPESGLOLRNKRIGTAAATLKLDPKVSSTNNLISITPSDNLAACTGN 919
      841 LHSQDMVFPESGLOLRNKRIGTAAATLKLDPKVSSTNNLISITPSDNLAACTGN 900
QY      920 TSSIGPSPMEVHYDSQDPTLFGKSSPLTESGSPILSEENNOSKLLSGLANSSQSSM 979
      901 TSSIGPSPMEVHYDSQDPTLFGKSSPLTESGSPILSEENNOSKLLSGLANSSQSSM 960
QY      980 GKNVSTESGRLEFKGRAGPALTLKNALFKYSISLTKNTSNKATNRKTHIGPSI 1039
      961 GKNVSTESGRLEFKGRAGPALTLKNALFKYSISLTKNTSNKATNRKTHIGPSI 1020
QY      1040 LIENSPVWONILIESDTEKKVTPLIHDRMLDKNATLALRLNHSNKTTSKKNEMVQOK 1099
      1021 LIENSPVWONILIESDTEKKVTPLIHDRMLDKNATLALRLNHSNKTTSKKNEMVQOK 1080
QY      1100 KEGPIPPDAQNPDMSEFKMLFLPESARMIQRTGKNSLNSGSGSPSKOLVSLGPEKSVEG 1159
      1081 KEGPIPPDAQNPDMSEFKMLFLPESARMIQRTGKNSLNSGSGSPSKOLVSLGPEKSVEG 1140
QY      1160 ONFLSEKNKYVVGKGFRTDVGLEKMPFSPSRMLFLTNLDNLHENHNOEKKTQOEIEK 1219
      1141 ONFLSEKNKYVVGKGFRTDVGLEKMPFSPSRMLFLTNLDNLHENHNOEKKTQOEIEK 1200
QY      1220 KETLIOENVVLPQIHTVGTGNFKMNLFLSTQNOYEGSYDGAVALPVLQDRSLNDSTNR 1279
      1201 KETLIOENVVLPQIHTVGTGNFKMNLFLSTQNOYEGSYDGAVALPVLQDRSLNDSTNR 1260
QY      1280 TKRHTAFHSKKEEENLEGLAQTKQIYEKYACTSITSPNTSQOIFYOQRKRALQOFR 1339
      1261 TKRHTAFHSKKEEENLEGLAQTKQIYEKYACTSITSPNTSQOIFYOQRKRALQOFR 1320
QY      1340 PLEETELEKRIIYDQISQWMSNMKHLTPSTQIOIYNKEKEGALTTQSPISDCLTRSHST 1399
      1321 PLEETELEKRIIYDQISQWMSNMKHLTPSTQIOIYNKEKEGALTTQSPISDCLTRSHST 1380
QY      1400 POANRSLPLIAKVSFSPSIRPIYTLRVLPDONSHPAASVYKKNOSGOESSHPLQCAKK 1459
      1381 POANRSLPLIAKVSFSPSIRPIYTLRVLPDONSHPAASVYKKNOSGOESSHPLQCAKK 1440
QY      1460 NNLSLALITLEMTGDOREYSGISGTSATNSVYKKEVNTVLPKPDLPKTSQVETLLPVRHT 1519
      1441 NNLSLALITLEMTGDOREYSGISGTSATNSVYKKEVNTVLPKPDLPKTSQVETLLPVRHT 1500
QY      1520 YQKDLFPTETSNQSPGHLDLVGSLQGTBGAIKWNEARPKGVFLVATESSAKTPSK 1579
      1501 YQKDLFPTETSNQSPGHLDLVGSLQGTBGAIKWNEARPKGVFLVATESSAKTPSK 1560
QY      1580 LLDPLANDNHYGTQIPEKEMKSOEKSPEKTAFFKKDDIILSLACSNNAIAINEGONRP 1639
      1561 LLDPLANDNHYGTQIPEKEMKSOEKSPEKTAFFKKDDIILSLACSNNAIAINEGONRP 1620
QY      1640 EIEYTAQKQRTRELCSQNPVYLRHOREITRTTISDOEEIDYDITSVYMKKEDEPDIT 1699
      1621 EIEYTAQKQRTRELCSQNPVYLRHOREITRTTISDOEEIDYDITSVYMKKEDEPDIT 1680
QY      1700 DEDENOSPRFOKTRTHYFIAAVERLMDYGSSPVLNRAQSSVPQKRVVFOEFTD 1759
      1681 DEDENOSPRFOKTRTHYFIAAVERLMDYGSSPVLNRAQSSVPQKRVVFOEFTD 1740
QY      1760 GSFQPLRYRGELNEHGLGAPYIRAEVEDNIMYTFPNQASRPYSYSSLSIYEEDQOQGA 1819
      1741 GSFQPLRYRGELNEHGLGAPYIRAEVEDNIMYTFPNQASRPYSYSSLSIYEEDQOQGA 1800
QY      1820 EPRKNFYKPEETKYTKYKQOHMAPTDEDEPCAKAYVESDYULEDVSGLIGPLVCHT 1879
      1801 EPRKNFYKPEETKYTKYKQOHMAPTDEDEPCAKAYVESDYULEDVSGLIGPLVCHT 1860

```

```

QY 1880 NTLPAGROVYVOEALFTIEDERKSWYTEMNERCRAPCIOIMEDPTFEKENTRHA 1939
DB 1861 NTLPAGROVYVOEALFTIEDERKSWYTEMNERCRAPCIOIMEDPTFEKENTRHA 1920
QY 1940 INGYIMDLPLGLVMAQDRIKMWILLSGNSNENIHSIHSFGHVTFRKEEYKMAIYNLP 1999
DB 1921 INGYIMDLPLGLVMAQDRIKMWILLSGNSNENIHSIHSFGHVTFRKEEYKMAIYNLP 1980
QY 2000 GVEFEVEMLPKSKAGIWRRECELGHEHLHAGKSTLFLVYSNKOQTPLGKASGHIDPQTAS 2059
DB 1981 GVEFEVEMLPKSKAGIWRRECELGHEHLHAGKSTLFLVYSNKOQTPLGKASGHIDPQTAS 2040
QY 2060 GYCGWAPKRLARLHSGSINNAKSTKEPFSKIVYDLAIIHIGIKTGAOKKSSLYTSQ 2119
DB 2041 GYCGWAPKRLARLHSGSINNAKSTKEPFSKIVYDLAIIHIGIKTGAOKKSSLYTSQ 2100
QY 2120 FIIMYSLDGKRWQYRGNSTGTLMVFFGNVDSGKIHNIENPPIIARYIRLHPTHSIRS 2179
DB 2101 FIIMYSLDGKRWQYRGNSTGTLMVFFGNVDSGKIHNIENPPIIARYIRLHPTHSIRS 2160
QY 2180 TLMELMGCDLNSGSMPLGMSKASIDAOITASSYFTNMFATNSPSKARLHLOGRSNMR 2239
DB 2161 TLMELMGCDLNSGSMPLGMSKASIDAOITASSYFTNMFATNSPSKARLHLOGRSNMR 2220
QY 2240 PUVNPKREMLQVDFPKTKYTGVTGQVKSLLFSMTYKEFLISSODGHOWTLFPONGKY 2299
DB 2221 PUVNPKREMLQVDFPKTKYTGVTGQVKSLLFSMTYKEFLISSODGHOWTLFPONGKY 2280
QY 2300 KVEQGNQDSFTPVVNSLDPPLTRYLRIHPOSVMHQIALRMEVAGCEADOLY 2351
DB 2281 KVEQGNQDSFTPVVNSLDPPLTRYLRIHPOSVMHQIALRMEVAGCEADOLY 2332

RESULT 10
; Sequence 2, Application US/09037601
; Patent No. 6180371
; GENERAL INFORMATION:
; APPLICANT: Lollar, John S.
; TITLE OF INVENTION: Hybrid Human/Animal Factor VIII
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenlee, Wmner and Sullivan, P.C.
; STREET: 5370 Manhattan Circle Suite 201
; CITY: Boulder
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/037,601
; FILING DATE: 26-JUN-1996
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/US94/13200
; FILING DATE: 15-NOV-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/212,133
; FILING DATE: 11-MAR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/864,004
; FILING DATE: 07-APR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Ferber, Donna M.
; REGISTRATION NUMBER: 33,878
; REFERENCE/DOCKET NUMBER: 75-95F
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 303/499-8080
; TELEFAX: 303/499-8089

```

```

; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2332 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; HYPOTHEICAL: YES
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; TISSUE TYPE: Liver
; US-09-037-601-2

Query Match          99.1%; Score 12301; DB 4; Length 2332;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2330; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 20 ATRRYVIGAVELSMQSDYQSDLGELPVDARPPRPVKSPPPTSVYKKTFLVEPTDHLFN 79
DB 1 ATRRYVIGAVELSMQSDYQSDLGELPVDARPPRPVKSPPPTSVYKKTFLVEPTDHLFN 60
QY 80 IAKRPBMGLGPTQAEYDVTVTTLNMAKSHVSLHAGVSYMKASEGAEYDQTSQ 139
DB 61 IAKRPBMGLGPTQAEYDVTVTTLNMAKSHVSLHAGVSYMKASEGAEYDQTSQ 120
QY 140 REKEDKVPFGSGSHYVQVYLKENGPMASDPLCTLYSYLHVLDLVKDLNSGLIGALLVCR 199
DB 121 REKEDKVPFGSGSHYVQVYLKENGPMASDPLCTLYSYLHVLDLVKDLNSGLIGALLVCR 180
QY 200 EGSIAKEKQTLHPTLIPVPEGKSMSEETKNSLMODDPAASARAPKMTGVGYNR 259
DB 181 EGSIAKEKQTLHPTLIPVPEGKSMSEETKNSLMODDPAASARAPKMTGVGYNR 240
QY 260 SLPLIGCHRSVYHVAIIGMTPPEVHSIFLEGHTLVNHRQASLEISPTFLTAQTL 319
DB 241 SLPLIGCHRSVYHVAIIGMTPPEVHSIFLEGHTLVNHRQASLEISPTFLTAQTL 300
QY 320 MDIGPFLFCJHISHQHGMETAVKYVDSQPEEPOLRMKNEEADYDDDLTQSEMDVYRF 379
DB 301 MDIGPFLFCJHISHQHGMETAVKYVDSQPEEPOLRMKNEEADYDDDLTQSEMDVYRF 360
QY 380 DDNSPSFTQIRSVAKKHPKTYWYHIAAEEDMDAPLYAPDDBSYKQYANRPOKIG 439
DB 361 DDNSPSFTQIRSVAKKHPKTYWYHIAAEEDMDAPLYAPDDBSYKQYANRPOKIG 420
QY 440 RYKRVKPMAYTDEFTKTRBAIQHESGILGPLLYGEVQDILLIIFKNAQSRPNITYPGI 499
DB 421 RYKRVKPMAYTDEFTKTRBAIQHESGILGPLLYGEVQDILLIIFKNAQSRPNITYPGI 480
QY 500 TDVRLPYSHRLPKGVKHLKDFPLIPGEIRKYKWTYVDDGPTKSDRCRLTRYSSFYVME 559
DB 481 TDVRLPYSHRLPKGVKHLKDFPLIPGEIRKYKWTYVDDGPTKSDRCRLTRYSSFYVME 540
QY 560 RDLASGLIGPLIICYKESVORONQIMSKRVILFSPDENSTVLTENORELPNAG 619
DB 541 RDLASGLIGPLIICYKESVORONQIMSKRVILFSPDENSTVLTENORELPNAG 600
QY 620 VQLEDEEFOASINHSINGYVDSIQLSVCLAEVAYWYIISIGAQDPLSVFSGYTRKH 679
DB 601 VQLEDEEFOASINHSINGYVDSIQLSVCLAEVAYWYIISIGAQDPLSVFSGYTRKH 660
QY 680 KMYEDTLTLPFSGGEVTFVMSMENPGLMTILGCHNSDFNGMGTALLKYSSCKKNGDYFE 739
DB 661 KMYEDTLTLPFSGGEVTFVMSMENPGLMTILGCHNSDFNGMGTALLKYSSCKKNGDYFE 720
QY 740 DSYEDISAVLISKNNAIETPSSQNSRSHPTROKOFNAITIPENDIEKTDPPFAHRTMP 799
DB 721 DSYEDISAVLISKNNAIETPSSQNSRSHPTROKOFNAITIPENDIEKTDPPFAHRTMP 780
QY 800 KIQNVSSDLMLTLROGPTPHGSLSDLOEAKYETFSDDSPALISNNSLSEMHFEPQ 859
DB 781 KIQNVSSDLMLTLROGPTPHGSLSDLOEAKYETFSDDSPALISNNSLSEMHFEPQ 770

```

Db 781 KIONVSSDILLMLRQSPTRHGLSLDQEAKEYETSDPSPGALDSNNLSXETHRPQ 840
Qy LHHGDMVFTPESSGLOLRINEKLTGTAATELKIDFVYSTSNLISITPBDNLAACTDN 919
Db LHHGDMVFTPESSGLOLRINEKLTGTAATELKIDFVYSTSNLISITPBDNLAACTDN 900
Qy 920 TSSLGPPSMVHYDSQLOTTLEFKKSSPLTESGGPLSLSENNDSKLESGLMSQSSW 979
Db 901 TSSLGPPSMVHYDSQLOTTLEFKKSSPLTESGGPLSLSENNDSKLESGLMSQSSW 960
Qy 980 GKNVSTESGRLFKGRAGPALTLTKDNALFKVSIISLTKNTKSNNSATNKTHTIDPSTL 1039
Db 961 GKNVSTESGRLFKGRAGPALTLTKDNALFKVSIISLTKNTKSNNSATNKTHTIDPSTL 1020
Qy 1040 LIENSPYMONILLESOTPEKKYTPILHDBMLMDKNATALLRNHMSKTTSSKNEMVQOK 1099
Db 1021 LIENSPYMONILLESOTPEKKYTPILHDBMLMDKNATALLRNHMSKTTSSKNEMVQOK 1080
Qy 1100 KEGPIPDQONPDMSEFKMLFPESARMIORTHGKNSLNSGQSPSKOLVSLGPEKSEVG 1159
Db 1081 KEGPIPDQONPDMSEFKMLFPESARMIORTHGKNSLNSGQSPSKOLVSLGPEKSEVG 1140
Qy 1160 ONFLSEKKNVYGCEFTKDVGLKEMVPESSRNLFNLNDNLHENTHNOEKKIOEIEK 1219
Db 1141 ONFLSEKKNVYGCEFTKDVGLKEMVPESSRNLFNLNDNLHENTHNOEKKIOEIEK 1200
Qy 1220 KETLIOENVYLPOLHVTGKXPMKMLFLSTRONVGSYDQAYAPVLODRSLNDSTNR 1279
Db 1201 KETLIOENVYLPOLHVTGKXPMKMLFLSTRONVGSYDQAYAPVLODRSLNDSTNR 1260
Qy 1280 TKKTHAFHSKKEEENLEGLNQTOIVEKYACTTRISPTSQONFVTORSKRALKOFRL 1339
Db 1261 TKKTHAFHSKKEEENLEGLNQTOIVEKYACTTRISPTSQONFVTORSKRALKOFRL 1320
Qy 1340 PLEETELEKRIYVDSTOWSKNMKHLTPSTLIOIDNEKEKALIOSPLSCLTGRSHSI 1399
Db 1321 PLEETELEKRIYVDSTOWSKNMKHLTPSTLIOIDNEKEKALIOSPLSCLTGRSHSI 1380
Qy 1400 POANRSLPLAKVSSPSIRPIYLRVLPONSSHRLAAYRKSQSGVOSHPLQCAKK 1459
Db 1381 POANRSLPLAKVSSPSIRPIYLRVLPONSSHRLAAYRKSQSGVOSHPLQCAKK 1440
Qy 1460 NNLSTAILTLEMTGDOREVSIGTSAINSVYKKEVNTVLPKPDLPKTSIGVYELLPVYHI 1519
Db 1441 NNLSTAILTLEMTGDOREVSIGTSAINSVYKKEVNTVLPKPDLPKTSIGVYELLPVYHI 1500
Qy 1520 YOKDLFPEYETSNCSGPHLDLYEBSLLQGTGEGAIKWNBANRPGRVPLRVATESSAKTPSK 1579
Db 1501 YOKDLFPEYETSNCSGPHLDLYEBSLLQGTGEGAIKWNBANRPGRVPLRVATESSAKTPSK 1560
Qy 1580 LLDPLAMNHYGTOIPKEEKSQEKSPKRTAFKKKDTILSLNACESNHALIAINQONKP 1639
Db 1561 LLDPLAMNHYGTOIPKEEKSQEKSPKRTAFKKKDTILSLNACESNHALIAINQONKP 1620
Qy 1640 EIEVTWAKQRTERLCSQNPVYLAVERLMDYGSSSPHVLRRAQSGSVQFKKVFQOETFD 1759
Db 1621 EIEVTWAKQRTERLCSQNPVYLAVERLMDYGSSSPHVLRRAQSGSVQFKKVFQOETFD 1680
Qy 1700 DEDENQSPRSFQKTRHYFLAVERLMDYGSSSPHVLRRAQSGSVQFKKVFQOETFD 1759
Db 1681 DEDENQSPRSFQKTRHYFLAVERLMDYGSSSPHVLRRAQSGSVQFKKVFQOETFD 1740
Qy 1760 GSFOTPLRGELNENHGLGPRYIRAVEEDNIMVFRNQRAPVSPFSSILSIEEORQGA 1819
Db 1741 GSFOTPLRGELNENHGLGPRYIRAVEEDNIMVFRNQRAPVSPFSSILSIEEORQGA 1800
Qy 1820 EPRKAPVAPNETKTYFKKVOHNAAPKQDFQCKAMAFSDVDLEKOVHSGLIGPLVCHT 1879
Db 1801 EPRKAPVAPNETKTYFKKVOHNAAPKQDFQCKAMAFSDVDLEKOVHSGLIGPLVCHT 1860
Qy 1880 NTLNPAHGRQYTVQEFALFTTIPDETCKSVYETENNERCAPCNIOEMDPTEKENVREHA 1939
Db 1861 NTLNPAHGRQYTVQEFALFTTIPDETCKSVYETENNERCAPCNIOEMDPTEKENVREHA 1920

Qy 1940 INGYIMDTLPGVMAODORIRWYLLSMGSENHITSHSGHVTYRREKEYKALYNLP 1999
Db 1921 INGYIMDTLPGVMAODORIRWYLLSMGSENHITSHSGHVTYRREKEYKALYNLP 1980
Qy 2000 GVFEVEMLPKSKAGIWRVECLIGEHLNAGSTFLVYSNRCOTPLGMAASHIRDPQITAS 2059
Db 1981 GVFEVEMLPKSKAGIWRVECLIGEHLNAGSTFLVYSNRCOTPLGMAASHIRDPQITAS 2040
Qy 2060 GYOQMAPKLARLHYSGSINASTREPPSWIKTDLAPMIRIGITOGAQRKSSLYISQ 2119
Db 2041 GYOQMAPKLARLHYSGSINASTREPPSWIKTDLAPMIRIGITOGAQRKSSLYISQ 2100
Qy 2120 FLIMYSIDGKKNOTYNGNSTGIMAFYGCNDSSGITHNFPNPIIARIYLRHPHYSIS 2179
Db 2101 FLIMYSIDGKKNOTYNGNSTGIMAFYGCNDSSGITHNFPNPIIARIYLRHPHYSIS 2160
Qy 2180 TLRNMLGCDLNSCNPGLMESKALISDAQITASSYFTNMFATWSPSKARLHLQGRSNMR 2239
Db 2161 TLRNMLGCDLNSCNPGLMESKALISDAQITASSYFTNMFATWSPSKARLHLQGRSNMR 2220
Qy 2240 POVNNPKEMLOVDFOKTMKVTGVTQGVKSLTSMYKKEFLISSQDGHQWTLFPQNGKY 2299
Db 2221 POVNNPKEMLOVDFOKTMKVTGVTQGVKSLTSMYKKEFLISSQDGHQWTLFPQNGKY 2280
Qy 2300 KYEGQNDSTPYVNSLDPDLTTRYLRIHPQSWYHOLALRMVTLCEAODLY 2351
Db 2281 KYEGQNDSTPYVNSLDPDLTTRYLRIHPQSWYHOLALRMVTLCEAODLY 2332

RESULT 11
US-09-315-179-2
Sequence 2, Application US/09315179
Patent No. 6376463
GENERAL INFORMATION:
APPLICANT: Lollar, John S
TITLE OF INVENTION: Modified Factor VIII
FILE REFERENCE: 75-95H
CURRENT APPLICATION NUMBER: US/09/315, 179
EARLIER FILING DATE: 1999-05-20
EARLIER APPLICATION NUMBER: U.S. 09/037, 601
EARLIER FILING DATE: 1998-03-10
EARLIER APPLICATION NUMBER: U.S. 08/670, 707
EARLIER FILING DATE: 1996-06-26
EARLIER APPLICATION NUMBER: PCT/US97/11155
EARLIER FILING DATE: 1997-06-26
EARLIER APPLICATION NUMBER: PCT/US94/13200
EARLIER FILING DATE: 1994-11-15
EARLIER APPLICATION NUMBER: U.S. 08/212, 133
EARLIER FILING DATE: 1994-03-11
EARLIER APPLICATION NUMBER: U.S. 07/864, 004
NUMBER OF SEQ ID NOS: 40
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 2
LENGTH: 2332
TYPE: PRT
ORGANISM: Homo sapiens
us-09-315-179-2

Query Match 99.1%; Score 12301; DB 4; Length 2332;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2330; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 20 ARRRYTGAVELSDMOSDLELVNDAPRRPKSFPRTNSVYKKTTLVEETHLEN 79
Db 1 AKRRYTGAVELSDMOSDLELVNDAPRRPKSFPRTNSVYKKTTLVEETHLEN 60
Qy 80 IAKRPPPMGLIGTTIOAEVYDTVYITLKNASHPVSLAAGVSYWKASDEGAEDDQTSQ 139
Db 61 IAKRPPPMGLIGTTIOAEVYDTVYITLKNASHPVSLAAGVSYWKASDEGAEDDQTSQ 120
Qy 140 RKEEDKVFPGCSHTYVMQVULKENPMASDCLTYSYLSHVDLVKDNLSLIGALLVCH 199

Db	121	REKEDKVEFGSGSHYVWQVLKENGPAASDPLCLVYLYSHYSHDVLKDLGSLGALLVCR	180	
Qy	200	EGSLANEKTOULHKEILLFAVEDEGKSMHSETKNSLMODRDAASRAMPKMHTVNGYVR	259	
Db	181	EGSLAKEKTOULHKEILLFAVEDEGKSMHSETKNSLMODRDAASRAMPKMHTVNGYVR	240	
Qy	260	SLPGLGCHRSVYMHVIGMGTPEVHSIFLEGHFLVYNHROASLEISPTFLAQTL	319	
Db	241	SLPGLGCHRSVYMHVIGMGTPEVHSIFLEGHFLVYNHROASLEISPTFLAQTL	300	
Qy	320	MDLQGLFLCHISSQHOGMEAYKYDUGCEFEQLRMKNNEAEYDDDLTDSMDVYVR	379	
Db	301	MDLQGLFLCHISSQHOGMEAYKYDUGCEFEQLRMKNNEAEYDDDLTDSMDVYVR	360	
Qy	380	DDNSPSFIOIRSVAKKHPKTWVHYIAEBEDMDYAPVLAPDDRKYKSOYLNNQFORIG	439	
Db	361	DDNSPSFIOIRSVAKKHPKTWVHYIAEBEDMDYAPVLAPDDRKYKSOYLNNQFORIG	420	
Qy	440	RKRYKRFMAAYTDEFTKTRREALIOHESGILGPLLYGEVSDTLILIFKNQASPRYNTYPHGI	499	
Db	421	RKRYKRFMAAYTDEFTKTRREALIOHESGILGPLLYGEVSDTLILIFKNQASPRYNTYPHGI	480	
Qy	500	TDVRLYSRRLPKGYKHLKDFPILPGSIFRYKMTVTVEDGPTKSDPCLTRYSSFYVNE	559	
Db	481	TDVRLYSRRLPKGYKHLKDFPILPGSIFRYKMTVTVEDGPTKSDPCLTRYSSFYVNE	540	
Qy	560	RDLASGLIGPLICYKESVDORGNOIMSDRNVILFSVEDERSWYLTENIORELPNPAQ	619	
Db	541	RDLASGLIGPLICYKESVDORGNOIMSDRNVILFSVEDERSWYLTENIORELPNPAQ	600	
Qy	620	VQLEDPFOASNMHSINGVYFDSLOJSCVGHVAVYTLISGATODLSTFSGSTFHR	679	
Db	601	VQLEDPFOASNMHSINGVYFDSLOJSCVGHVAVYTLISGATODLSTFSGSTFHR	660	
Qy	680	KMYEDTLILFPESGSEVYFVSMENPGMLIGCHNSDFRNRGMTALLKYSCKDKNTGDYIE	739	
Db	661	KMYEDTLILFPESGSEVYFVSMENPGMLIGCHNSDFRNRGMTALLKYSCKDKNTGDYIE	720	
Qy	740	DSYEDISAYLISKNNNAIEBSPFSONSRHPSTROKOFNATITPENDIEKTDWFHARTPM	799	
Db	721	DSYEDISAYLISKNNNAIEBSPFSONSRHPSTROKOFNATITPENDIEKTDWFHARTPM	780	
Qy	800	KIONVSSDMLMLRSPPHGLSISLOBAKETESDPSGALIDSNSISLSEMTHRRQ	859	
Db	781	KIONVSSDMLMLRSPPHGLSISLOBAKETESDPSGALIDSNSISLSEMTHRRQ	840	
Qy	860	LHSGDMVETPESGLOLRLNEKLGTTAATBLAKLDFKVSSTSNMLISTIPSDNLAAGTON	919	
Db	841	LHSGDMVETPESGLOLRLNEKLGTTAATBLAKLDFKVSSTSNMLISTIPSDNLAAGTON	900	
Qy	920	TSLGPPSMFVHYHSDQDLTTLFGKSSPLTESGSPILSEENNDKLLSEGLMNSOESSM	979	
Db	901	TSLGPPSMFVHYHSDQDLTTLFGKSSPLTESGSPILSEENNDKLLSEGLMNSOESSM	960	
Qy	980	GKNVSTESGRLEFKGRAGPALTLKONALFKYISILKTNKTSNNAKTHHIDPST	1039	
Db	961	GKNVSTESGRLEFKGRAGPALTLKONALFKYISILKTNKTSNNAKTHHIDPST	1020	
Qy	1040	LIENSPPWONILLESDETEKKTPLIHDRLMLDKNATLRLNHNKNTSSKNMEMYQOK	1099	
Db	1021	LIENSPPWONILLESDETEKKTPLIHDRLMLDKNATLRLNHNKNTSSKNMEMYQOK	1080	
Qy	1100	KEGP1PPDAONPMSFTKMLPLPESAWIORTHGKNSLNSGGSPKOLVSLGPEKSVBE	1159	
Db	1081	KEGP1PPDAONPMSFTKMLPLPESAWIORTHGKNSLNSGGSPKOLVSLGPEKSVBE	1140	
Qy	1160	ONFLSEKKNVYVGKGFETDVLKEMVPPSSRNILFLNLDNLEHNNNTNOKKIOEIERK	1219	
Db	1141	ONFLSEKKNVYVGKGFETDVLKEMVPPSSRNILFLNLDNLEHNNNTNOKKIOEIERK	1200	
Qy	1220	KETLIOENVYLPOIHVVGTKNPKMKNFLSTGONVSGSYDAVAPVLODRSLNOSTNR	1279	
Db	1201	KETLIOENVYLPOIHVVGTKNPKMKNFLSTGONVSGSYDAVAPVLODRSLNOSTNR	1260	
Qy	1280	TKRHTAHFSKGEENLEGLONOTKOIWEKTACTRISPNTSQONFYORSKALAKOFTL	1339	
Db	1261	TKRHTAHFSKGEENLEGLONOTKOIWEKTACTRISPNTSQONFYORSKALAKOFTL	1320	
Qy	1340	PLBETELERKRIIVDPTSPQMSKNNKHLPSTLQIDUYNEKKGALITOSPLSDCLTRSHSI	1399	
Db	1321	PLBETELERKRIIVDPTSPQMSKNNKHLPSTLQIDUYNEKKGALITOSPLSDCLTRSHSI	1380	
Qy	1400	POANRSPILAVSSPFSIRIYTLRPLRDNSSHLPAASRYKRDGVOSSSHFLOGAK	1459	
Db	1381	POANRSPILAVSSPFSIRIYTLRPLRDNSSHLPAASRYKRDGVOSSSHFLOGAK	1440	
Qy	1460	NULSLAILFLEMIGDQREVSGLSATNSVYKYKVENTVLPKPDLPYSGKVELLPKVI	1519	
Db	1441	NULSLAILFLEMIGDQREVSGLSATNSVYKYKVENTVLPKPDLPYSGKVELLPKVI	1500	
Qy	1520	YONDLPPTETNSGSPGHDLYEGSLQGTGALIKMNEANRGKVPFLRVATESSARTPSK	1579	
Db	1501	YONDLPPTETNSGSPGHDLYEGSLQGTGALIKMNEANRGKVPFLRVATESSARTPSK	1560	
Qy	1580	LDPFLANDHYGTQJPKREKKSQKSPKTAFFKKDTILSLNACESNHAIALINEQONP	1639	
Db	1561	LDPFLANDHYGTQJPKREKKSQKSPKTAFFKKDTILSLNACESNHAIALINEQONP	1620	
Qy	1640	EIEYTWAKOGRTERLCSNPVLRKHOREITRTTLOSDEEIDYDTISVEMKKEDEFDY	1699	
Db	1621	EIEYTWAKOGRTERLCSNPVLRKHOREITRTTLOSDEEIDYDTISVEMKKEDEFDY	1680	
Qy	1700	DEDENOSPFRFOKTRHFIYAVERLDWYMSSSPVLRNROSQSVPOKRYVPOEFD	1759	
Db	1681	DEDENOSPFRFOKTRHFIYAVERLDWYMSSSPVLRNROSQSVPOKRYVPOEFD	1740	
Qy	1760	GSTFOPLIRGELNEHLGLGPIYIAEVEDINMTFPOASRYSFYSLSIYEDDROQA	1819	
Db	1741	GSTFOPLIRGELNEHLGLGPIYIAEVEDINMTFPOASRYSFYSLSIYEDDROQA	1800	
Qy	1820	EPKRNFKPNETKTYFMKVOHMAPTDOEEDCKAMAYFSDVLEKDVHSGILGPLLYCHT	1879	
Db	1801	EPKRNFKPNETKTYFMKVOHMAPTDOEEDCKAMAYFSDVLEKDVHSGILGPLLYCHT	1860	
Qy	1880	NTLNPARGOVYVOEALFETIPDETSMVTFENMRNCRACNOMEDDPFKENRPHA	1939	
Db	1861	NTLNPARGOVYVOEALFETIPDETSMVTFENMRNCRACNOMEDDPFKENRPHA	1920	
Qy	1940	INGYIMDTLPGVMAODORIRMYLILSMGSNENHSHIFSGHYFYRKKEEKALYNLTP	1999	
Db	1921	INGYIMDTLPGVMAODORIRMYLILSMGSNENHSHIFSGHYFYRKKEEKALYNLTP	1980	
Qy	2000	GVEFTEVEMLPSKAGIWRVECLIGELHLAGMSTFLVYSNKCQTPLGMAHGHIDFOITAS	2059	
Db	1981	GVEFTEVEMLPSKAGIWRVECLIGELHLAGMSTFLVYSNKCQTPLGMAHGHIDFOITAS	2040	
Qy	2060	GQYQWAPKILARLHYSGSTINAMSTKEPSPKIXDLAPINHIGITQOARCKSSLYIQ	2119	
Db	2041	GQYQWAPKILARLHYSGSTINAMSTKEPSPKIXDLAPINHIGITQOARCKSSLYIQ	2100	
Qy	2120	FIWYSIDGKKMOTYKNGSTGLWVFEFGVNDSSGKIHNFNPPIIARYIRLHPHYHS	2179	
Db	2101	FIWYSIDGKKMOTYKNGSTGLWVFEFGVNDSSGKIHNFNPPIIARYIRLHPHYHS	2160	
Qy	2180	TLRHELMGOLNSGSPPLGMSKALISNOATASVYTNMFAWMSKARLHIGORSNAR	2239	
Db	2161	TLRHELMGOLNSGSPPLGMSKALISNOATASVYTNMFAWMSKARLHIGORSNAR	2220	
Qy	2240	POVNNPKEMWLDVDFOKTKMYKVTGVTQGVKSILTSMTYKBEFLISSODGHOVTLFPONGY	2299	
Db	2221	POVNNPKEMWLDVDFOKTKMYKVTGVTQGVKSILTSMTYKBEFLISSODGHOVTLFPONGY	2280	
Qy	2300	KVFOGNDSTFPVYNSLDPPLTLRILRHQSVWVHIOALRMEVLCSEADODY	2351	
Db	2281	KVFOGNDSTFPVYNSLDPPLTLRILRHQSVWVHIOALRMEVLCSEADODY	2332	

```
RESULT 12
US-09-523-656-2
; Sequence 2, Application us/09523656
; Patent No. 6458563
; GENERAL INFORMATION:
; APPLICANT: Lollar S., John
; TITLE OF INVENTION: MODIFIED FACTOR VIII
; FILE REFERENCE: 75-951
; CURRENT APPLICATION NUMBER: US/09/523, 656
; EARLIER FILING DATE: 2000-03-10
; EARLIER APPLICATION NUMBER: 09/037, 601
; EARLIER FILING DATE: 1998-03-10
; EARLIER APPLICATION NUMBER: 08/670, 707
; EARLIER FILING DATE: 1996-06-26
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO: 2
; LENGTH: 2332
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-523-656-2

Query Match          99.1%; Score 12301; DB 4; Length 2332;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2330; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      20  ATRRYVIGAVELSDMDQSDGLGELPYDARPPRPKSPFPNTSVYKKTLFVEFTDLFN 79
DB      1  ATRRYVIGAVELSDMDQSDGLGELPYDARPPRPKSPFPNTSVYKKTLFVEFTDLFN 60
QY      80  IAKRPFPWGLGLPTIOAEVDTVTYTLKNASHPVSLHAGVSTWKASEGAEYDQTSQ 139
DB      61  IAKRPFPWGLGLPTIOAEVDTVTYTLKNASHPVSLHAGVSTWKASEGAEYDQTSQ 120
QY      140 REKEDRVFPFGSHTYVMQVLKENGPMASDPLCTYSYLSHYDLVKDLSGLIGALVCR 199
DB      121 REKEDRVFPFGSHTYVMQVLKENGPMASDPLCTYSYLSHYDLVKDLSGLIGALVCR 180
QY      200 EGSIAKERTQTLHKFTLFAVDEGKSHSTKNSLMODRDAASARPKMHTYNGVYR 259
DB      181 EGSIAKERTQTLHKFTLFAVDEGKSHSTKNSLMODRDAASARPKMHTYNGVYR 240
QY      260 SLPLGICHRKSYTWKHYVIGMGTTPVNSIFLEGHTFLVRNHQASLEISPLPTTAQTL 319
DB      241 SLPLGICHRKSYTWKHYVIGMGTTPVNSIFLEGHTFLVRNHQASLEISPLPTTAQTL 300
QY      320 MDLQFLFCHISSHODMEAYVVDSCPEEPQLRKNNNEBAEDYDDDLTDSMDVYRF 379
DB      301 MDLQFLFCHISSHODMEAYVVDSCPEEPQLRKNNNEBAEDYDDDLTDSMDVYRF 360
QY      380 DDDNSPFIQIRSVAKKHKPTWVHYIAEEDMDVAPVLAPODSSYKSOYLNNGPORIG 439
DB      361 DDDNSPFIQIRSVAKKHKPTWVHYIAEEDMDVAPVLAPODSSYKSOYLNNGPORIG 420
QY      440 RKYKRVPMAYTDEFTKTRREALIOHESGILGPLLYGEGDTLLIFRKNQASRPYIYHGI 499
DB      421 RKYKRVPMAYTDEFTKTRREALIOHESGILGPLLYGEGDTLLIFRKNQASRPYIYHGI 480
QY      500 TDVRLYSRRLPKGVNHLKDFPLPGELFFKMYVVEGDPKSPRCLTYTSSPFNME 559
DB      481 TDVRLYSRRLPKGVNHLKDFPLPGELFFKMYVVEGDPKSPRCLTYTSSPFNME 540
QY      560 RDLASGLIGPLLICYESVDQNGNOIMSKDNVYLSYVDENRNSYLTENIQRLPMPAG 619
DB      541 RDLASGLIGPLLICYESVDQNGNOIMSKDNVYLSYVDENRNSYLTENIQRLPMPAG 600
QY      620 VQLEDPFEOASNTMHSINGYVDSIQSLCHEVAAYWYILSIGAOTDFLSVFESGYTFKH 679
DB      601 VQLEDPFEOASNTMHSINGYVDSIQSLCHEVAAYWYILSIGAOTDFLSVFESGYTFKH 660
QY      680 KMVYEDTLTLFPFSGETVMSMENPGMIILGCHNSDFRNRGMTALLKVSQCDKWTGYYE 729
```

```
DB      661 KMVYEDTLTLFPFSGETVMSMENPGMIILGCHNSDFRNRGMTALLKVSQCDKWTGYYE 720
QY      740 DSYEDISAVYLLSKNNAIEPRSPQNSRHPSTROKQNPATITPNDIEKTDWPFNHRTPMP 799
DB      721 DSYEDISAVYLLSKNNAIEPRSPQNSRHPSTROKQNPATITPNDIEKTDWPFNHRTPMP 780
QY      800 KIQWSSSDILMLLRSPPHGLISLQDAKAKETSDPSGALDSNNSLSEMTFHRQ 859
DB      781 KIQWSSSDILMLLRSPPHGLISLQDAKAKETSDPSGALDSNNSLSEMTFHRQ 840
QY      860 LHSQDVFYFPESGILQRLINEKLGTAATEKIKLDFRVSSTNNLISIPSDNIAQTDN 919
DB      841 LHSQDVFYFPESGILQRLINEKLGTAATEKIKLDFRVSSTNNLISIPSDNIAQTDN 900
QY      920 TSSGPPSMVHYDQDQDTTLFGKSSPLTESGAPLSSEENNDKLLSEGLMNSQESSM 979
DB      901 TSSGPPSMVHYDQDQDTTLFGKSSPLTESGAPLSSEENNDKLLSEGLMNSQESSM 960
QY      980 GRNVSTESGRLEKGRAGPALLTKONALPKYISILTKNTSNNSATNKRTHIDGPSL 1039
DB      961 GRNVSTESGRLEKGRAGPALLTKONALPKYISILTKNTSNNSATNKRTHIDGPSL 1020
QY      1040 LIENSPWONILSDTEFEKVTPLIHDRMLMDKNATLRLNHSNKTSSKNNENVOQK 1099
DB      1021 LIENSPWONILSDTEFEKVTPLIHDRMLMDKNATLRLNHSNKTSSKNNENVOQK 1080
QY      1100 KEGPIPPDAQNPMSFVKMLFPESARMIORTGKNSLNSGGPSPKOLVSLGPEKSEVG 1159
DB      1081 KEGPIPPDAQNPMSFVKMLFPESARMIORTGKNSLNSGGPSPKOLVSLGPEKSEVG 1140
QY      1160 QNFISEKNKYVVGKGFYTKDQGLAKMFPSPSRNLPLTNLDNHENNTHQOKKIOEBIEK 1219
DB      1141 QNFISEKNKYVVGKGFYTKDQGLAKMFPSPSRNLPLTNLDNHENNTHQOKKIOEBIEK 1200
QY      1220 KETLIOENVVLPQIRHVNTGKTFKFNKNTFLSTRONVSGSYDAVAPVLQDRSLNDSTNR 1279
DB      1201 KETLIOENVVLPQIRHVNTGKTFKFNKNTFLSTRONVSGSYDAVAPVLQDRSLNDSTNR 1260
QY      1280 TKKHTAHFSKKGEEENLEGLQNOTQIYEKACTTRISPTSONFVYORSRALKOFR 1339
DB      1261 TKKHTAHFSKKGEEENLEGLQNOTQIYEKACTTRISPTSONFVYORSRALKOFR 1320
QY      1340 PLEETEELKRIYVDSTQWKNKMLTSPSTLQIDYNEKKGATOSPDLCTLRSHS1 1399
DB      1321 PLEETEELKRIYVDSTQWKNKMLTSPSTLQIDYNEKKGATOSPDLCTLRSHS1 1380
QY      1400 POANRSPPLIAKYSFSPSIRPYLTRVLFQDNSSHLPAASYRKKDSGYOESSHFLQGA 1459
DB      1381 POANRSPPLIAKYSFSPSIRPYLTRVLFQDNSSHLPAASYRKKDSGYOESSHFLQGA 1440
QY      1460 NNLSTATITLTEMGDQREVGSIGTSATNSYTKKVENTVLPKPDLPKTSIGVELLPVHI 1519
DB      1441 NNLSTATITLTEMGDQREVGSIGTSATNSYTKKVENTVLPKPDLPKTSIGVELLPVHI 1500
QY      1520 YOKDLFLETNSGSPCHLDLVESLSLQTBGAIKWNENRPGKVPPLVATRESSAKTPSK 1579
DB      1501 YOKDLFLETNSGSPCHLDLVESLSLQTBGAIKWNENRPGKVPPLVATRESSAKTPSK 1560
QY      1580 LLDPLAMDNHNGTQIKREEMKSOEKSPEKTKYFKKDTIILSNACSNHAIINEGONKP 1639
DB      1561 LLDPLAMDNHNGTQIKREEMKSOEKSPEKTKYFKKDTIILSNACSNHAIINEGONKP 1620
QY      1640 ELEVYMAKQKTERLCSQNPVLAKRHOEITRTTLQSDDEIDYDTISVEMKKEDDYI 1699
DB      1621 ELEVYMAKQKTERLCSQNPVLAKRHOEITRTTLQSDDEIDYDTISVEMKKEDDYI 1680
QY      1700 DEDENQSPRSFORKTRHYVIAAVERLMDYGNSSPHVLRNRAQSGSVQFKKVVOEFTD 1759
DB      1681 DEDENQSPRSFORKTRHYVIAAVERLMDYGNSSPHVLRNRAQSGSVQFKKVVOEFTD 1740
QY      1760 GSFQOPLYRGELNEHGLGPYIRAEVNDNIMVFRQASRPYSYSSLIYEEDQOGA 1819
```



```

Db      1741 GSFTOPLYRGELNHELGLGPIYIAEVEDINWTRNOASPRFTYSLSISTEEDORGA 1800
QY      1820 EPRKNEFKPNETKTYFWKVOHHMAPTKDEEDCKAMAYFSVDLDEKDVHSLIGPLVCHT 1879
Db      1801 EPRKNEFKPNETKTYFWKVOHHMAPTKDEEDCKAMAYFSVDLDEKDVHSLIGPLVCHT 1860
QY      1880 NTLNPAHGRQVYQVEAFLEFTIFEDTSKMTFENNERCRAPIQIOMEDPTFEKENTRPA 1939
Db      1861 NTLNPAHGRQVYQVEAFLEFTIFEDTSKMTFENNERCRAPIQIOMEDPTFEKENTRPA 1920
QY      1940 INGYIMDTPLGLVMAQDRIKMTYLLSGNSNENHSHSGVFTVVRKEEYKALYNLP 1999
Db      1921 INGYIMDTPLGLVMAQDRIKMTYLLSGNSNENHSHSGVFTVVRKEEYKALYNLP 1980
QY      2000 GVEFEVEMLPKAGIMRVECLIGEHLAGNSTLFLVYSNKCQTPLEGASGHIRDQITAS 2059
Db      1981 GVEFEVEMLPKAGIMRVECLIGEHLAGNSTLFLVYSNKCQTPLEGASGHIRDQITAS 2040
QY      2060 GYGGWAPKLARLHYSGSINAMSTKEPFSWIKYDILAPMIIHGKIKTGAROKFSSLYISQ 2119
Db      2041 GYGGWAPKLARLHYSGSINAMSTKEPFSWIKYDILAPMIIHGKIKTGAROKFSSLYISQ 2100
QY      2120 FITYSLDGKRWQYTRGNSGTGLMVFPGVNDSSGIRKHNIFNPPIIARYIRLAPFTHSIRS 2179
Db      2101 FITYSLDGKRWQYTRGNSGTGLMVFPGVNDSSGIRKHNIFNPPIIARYIRLAPFTHSIRS 2160
QY      2180 TLMELMAGDLNCSMPKMGESKASIDPAQITASSYFTNMFAWSPSKARLHLQGRSNMR 2239
Db      2161 TLMELMAGDLNCSMPKMGESKASIDPAQITASSYFTNMFAWSPSKARLHLQGRSNMR 2220
QY      2240 POWNPKEMLQYDPKTKMYTGVTOGVSKSLTSMYVKEFLISSOGHOMTLEFPNGVY 2299
Db      2221 POWNPKEMLQYDPKTKMYTGVTOGVSKSLTSMYVKEFLISSOGHOMTLEFPNGVY 2280
QY      2300 KVFQGNDSFTPVVNSLDPLLRYLRIRHPOQVHOIALRMEVLGCEADOLY 2351
Db      2281 KVFQGNDSFTPVVNSLDPLLRYLRIRHPOQVHOIALRMEVLGCEADOLY 2332

```

RESULT 13

PCT-US93-03275-4
; Sequence 4, Application PC/TUS9303275

; GENERAL INFORMATION:

; APPLICANT: Lollar, John S.

; TITLE OF INVENTION: Hybrid Human/Porcine Factor VIII

; NUMBER OF SEQUENCES: 6

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Kilpatrick & Cody

; STREET: 1100 Peachtree Street, Suite 2800

; CITY: Atlanta

; STATE: Georgia

; COUNTRY: US

; ZIP: 30309-4530

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: PCT/US93/03275

; FILING DATE: 19930407

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/864004

; FILING DATE: 07-APR-1992

; ATTORNEY/AGENT INFORMATION:

; NAME: Padst, Patrea L.

; REGISTRATION NUMBER: 31,284

; REFERENCE/DOCKET NUMBER: EMU 106PCT

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 404-815-6508

; TELEFAX: 404-815-6555

; TELEPHONE: 404-815-6508

```

; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2332 amino acids
; TYPE: AMINO ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Homo sapien
; TISSUE TYPE: Liver cDNA sequence
PCT-US93-03275-4

Query Match          99.1%; Score 12301; DB 5; Length 2332;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2330; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      20 ATRRYTLGAVELSDMYNQSGLGELPVDARPPRRVRSKSPFTSVYKTLFVETDHLFN 79
Db      1 ATRRYTLGAVELSDMYNQSGLGELPVDARPPRRVRSKSPFTSVYKTLFVETDHLFN 60
QY      80 IARPRPMGGLGPTIQAEYDVYVITLKMAASHVSLHAGVSWKASGAEYDQTSQ 139
Db      61 IARPRPMGGLGPTIQAEYDVYVITLKMAASHVSLHAGVSWKASGAEYDQTSQ 120
QY      140 REKEDKVPFGSGSHYVQVYKENGPMASDPLCTAYSLSHVDLVKDLNSGLIGALVCR 199
Db      121 REKEDKVPFGSGSHYVQVYKENGPMASDPLCTAYSLSHVDLVKDLNSGLIGALVCR 180
QY      200 EGSIAEKQTQTLHAKTILFAVFDGKSMHSETRKNSLMODRDAASARMPMHTVNGYVNR 259
Db      181 EGSIAEKQTQTLHAKTILFAVFDGKSMHSETRKNSLMODRDAASARMPMHTVNGYVNR 240
QY      260 SLPGILGCHRRKSVYWHVYIGMGTPEVHSIFLGGHTFLVNRHQASLEISFTFLAQTL 319
Db      241 SLPGILGCHRRKSVYWHVYIGMGTPEVHSIFLGGHTFLVNRHQASLEISFTFLAQTL 300
QY      320 MDLGOFLFCHTSHSHOHGMAYVKTQVDSCEPPEPOLRMKNNEADYDDLTSEMDVYRF 379
Db      301 MDLGOFLFCHTSHSHOHGMAYVKTQVDSCEPPEPOLRMKNNEADYDDLTSEMDVYRF 360
QY      380 DDNSPSFTQIRSVAKKHPKTHWYIAAEEEDMAYPLVLAADDRSYKSOYLNGPQRTG 439
Db      361 DDNSPSFTQIRSVAKKHPKTHWYIAAEEEDMAYPLVLAADDRSYKSOYLNGPQRTG 420
QY      440 RYKKTFRMAAYTDEFKTRERAIQHESGILGPLLGEVGDTLITFFKNQASRPYNIYPIGI 499
Db      421 RYKKTFRMAAYTDEFKTRERAIQHESGILGPLLGEVGDTLITFFKNQASRPYNIYPIGI 480
QY      500 TDVRLXSRRLPKGYKHKDPPLIPGIRKRYKWTYVDEGRTKSDPCLTRYSSVYVME 559
Db      481 TDVRLXSRRLPKGYKHKDPPLIPGIRKRYKWTYVDEGRTKSDPCLTRYSSVYVME 540
QY      560 RDIASGLIGPLLICIKESVDQRNQIMSKRNVILSVFEDNRKWTLTENIQFLPNPAG 619
Db      541 RDIASGLIGPLLICIKESVDQRNQIMSKRNVILSVFEDNRKWTLTENIQFLPNPAG 600
QY      620 VOLEDPEFOASINMHSINGYFDSIQLSVCLHEVYVYVILSIGAQTFLSVFSGVTFFH 679
Db      601 VOLEDPEFOASINMHSINGYFDSIQLSVCLHEVYVYVILSIGAQTFLSVFSGVTFFH 660
QY      680 KMYEDTLTLFPFSGEYVPMSEKPGIMLIGCHNSDPFRNRKATLLAVSSCDKNTGDIYE 739
Db      661 KMYEDTLTLFPFSGEYVPMSEKPGIMLIGCHNSDPFRNRKATLLAVSSCDKNTGDIYE 720
QY      740 DSYEDISAYILSKNNAIEPRFSQNSRHSPTQKQFNATITPENDIKETPPMFAHRTPM 799
Db      721 DSYEDISAYILSKNNAIEPRFSQNSRHSPTQKQFNATITPENDIKETPPMFAHRTPM 780
QY      800 KIONWSSDLMMLRQSTPRGSLSDQEAKEYTFSDPBGALIDNNLSKMTIFRPQ 859

```

```

Db      781 KIONVSSDLMLLNQSPTRPHGLSLDLOEAXEYTFSDPSPGALDSDNSLSMETHFRBPQ 840
QY      860 LAHSGDMVTPSPSGLOLRLNEKIGTATATLKLDPKVSYSNNLISITSDMLAAGTGN 919
Db      841 LAHSGDMVTPSPSGLOLRLNEKIGTATATLKLDPKVSYSNNLISITSDMLAAGTGN 900
QY      920 TSSLGPPSPMHYDSDLTTLFGKSSPLTESSGSPLSSEENDSKLIEGSLANSSQSSW 979
Db      901 TSSLGPPSPMHYDSDLTTLFGKSSPLTESSGSPLSSEENDSKLIEGSLANSSQSSW 960
QY      980 GKNSSTESGRLFEKGRRAHPALLTKDNALFKVYSILKTKNTSNKATNKTTHIGPSEL 1039
Db      961 GKNSSTESGRLFEKGRRAHPALLTKDNALFKVYSILKTKNTSNKATNKTTHIGPSEL 1020
QY      1040 LIENSPVWONLIESPTEKRYPLIDHRLMKNATLRLHNSNNTSSKMEWAOOK 1099
Db      1021 LIENSPVWONLIESPTEKRYPLIDHRLMKNATLRLHNSNNTSSKMEWAOOK 1080
QY      1100 KEGPIPPDAONPDMSEFKMLPLPSAAMIORHGNLSNGGSPSKOLVSLGPERSVBG 1159
Db      1081 KEGPIPPDAONPDMSEFKMLPLPSAAMIORHGNLSNGGSPSKOLVSLGPERSVBG 1140
QY      1160 ONFLSEKNKVVYKGEFTDVKLEKEMVPSRRMLFTNLIDNLHENHNGEKKIOEIERK 1219
Db      1141 ONFLSEKNKVVYKGEFTDVKLEKEMVPSRRMLFTNLIDNLHENHNGEKKIOEIERK 1200
QY      1220 KETLIOENVVLPQIHVYVGTNKNPKNLFILSTQONWBSGDAPVLODFRSLNOSTNR 1279
Db      1201 KETLIOENVVLPQIHVYVGTNKNPKNLFILSTQONWBSGDAPVLODFRSLNOSTNR 1260
QY      1280 TKKHTAHFSKKEENLEGLNQTQOYERKACTTRISBNTSQONFTVORSKRALQOFRL 1339
Db      1261 TKKHTAHFSKKEENLEGLNQTQOYERKACTTRISBNTSQONFTVORSKRALQOFRL 1320
QY      1340 PLEETELEKRIIYDSTSTQMSKNMKHLPTSLTQIYNNKEKGAITQSPDCLTNSHSI 1399
Db      1321 PLEETELEKRIIYDSTSTQMSKNMKHLPTSLTQIYNNKEKGAITQSPDCLTNSHSI 1380
QY      1400 POANRSPPLIAKVSPPSIRPIYLRVLFDONSHLPAASYKRRKSGVQSSHFLGQAKK 1459
Db      1381 POANRSPPLIAKVSPPSIRPIYLRVLFDONSHLPAASYKRRKSGVQSSHFLGQAKK 1440
QY      1460 NNLSLAILLTLEMTGOREVGLSGTSAINSVTYKVENYVLPKPDLPKTSKAVELLPPVHI 1519
Db      1441 NNLSLAILLTLEMTGOREVGLSGTSAINSVTYKVENYVLPKPDLPKTSKAVELLPPVHI 1500
QY      1520 YOKDLFPTETSNQSPGHLDLVEGSLLOGTBGALKNWENARPGKVPDLVATESSAKTPSK 1579
Db      1501 YOKDLFPTETSNQSPGHLDLVEGSLLOGTBGALKNWENARPGKVPDLVATESSAKTPSK 1560
QY      1580 LLDPLANDNHGTQOIRKEEMKSOEKSPEKTAFFKKDITLILNACHSNALAIWEGONKP 1639
Db      1561 LLDPLANDNHGTQOIRKEEMKSOEKSPEKTAFFKKDITLILNACHSNALAIWEGONKP 1620
QY      1640 ELEVYMAKOGRTERLCSNPPVLAKEHOREITRTTLOSDEEIDYDTISVEMKEDFDIY 1699
Db      1621 ELEVYMAKOGRTERLCSNPPVLAKEHOREITRTTLOSDEEIDYDTISVEMKEDFDIY 1680
QY      1700 DEDENOSPRSFOKKTTHYFIAAVERLMDYGSSPHVLRNRAQSSGVQFKVVFQEFDT 1759
Db      1681 DEDENOSPRSFOKKTTHYFIAAVERLMDYGSSPHVLRNRAQSSGVQFKVVFQEFDT 1740
QY      1760 GSFTOPLYRGLNENHGLGPIYIRAEVYDNIMTFRNOASRPYSYSSLSIEEDQOQGA 1819
Db      1741 GSFTOPLYRGLNENHGLGPIYIRAEVYDNIMTFRNOASRPYSYSSLSIEEDQOQGA 1800
QY      1820 EPRKNFYKPNETKTYWKVQHMAPTKDEPCKAMAFESDVLDLEOVHSGILGLVYCHT 1879
Db      1801 EPRKNFYKPNETKTYWKVQHMAPTKDEPCKAMAFESDVLDLEOVHSGILGLVYCHT 1860
QY      1880 NTLNPAHQROYVQOEALFFTIIFDETSKSWYFTENMERNCACPNQIOMEDPTFKENYRFAH 1939
Db      1861 NTLNPAHQROYVQOEALFFTIIFDETSKSWYFTENMERNCACPNQIOMEDPTFKENYRFAH 1920

```

```

QY      1940 INGIYMDTLRGLVMAODRIRWYLLMSGNSNENHSHFSGHVTFRKKEEYKMALYNXP 1999
Db      1921 INGIYMDTLRGLVMAODRIRWYLLMSGNSNENHSHFSGHVTFRKKEEYKMALYNXP 1980
QY      2000 GVEFTEYMLPSRAGIWRVECLIGELHLAGKSTLFLVYSNKCOTPLGMAASHIRDOQTAS 2059
Db      1981 GVEFTEYMLPSRAGIWRVECLIGELHLAGKSTLFLVYSNKCOTPLGMAASHIRDOQTAS 2040
QY      2060 GQYGNAPKILARHSGSINAMSTKEPFWIKVDLAPMIHGIKTOGARQKFSLSYQ 2119
Db      2041 GQYGNAPKILARHSGSINAMSTKEPFWIKVDLAPMIHGIKTOGARQKFSLSYQ 2100
QY      2120 FIIMYSIDCKKNOTRGNSGTGLWVFGVNDSGIKHNIFNPILAIRYLHPTHSIRS 2179
Db      2101 FIIMYSIDCKKNOTRGNSGTGLWVFGVNDSGIKHNIFNPILAIRYLHPTHSIRS 2160
QY      2180 TLREMLGCDLNSCAMPJAMESKRAISDAQITASSYFTNNFATWSPSARLHLOGRSNMR 2239
Db      2161 TLREMLGCDLNSCAMPJAMESKRAISDAQITASSYFTNNFATWSPSARLHLOGRSNMR 2220
QY      2240 POUVNPKEMLQVDFQRTMKVYGVTTQGVKSILTSWYKFEFLISSODGHQWTLFQNGKV 2299
Db      2221 POUVNPKEMLQVDFQRTMKVYGVTTQGVKSILTSWYKFEFLISSODGHQWTLFQNGKV 2280
QY      2300 KVFQGNODSEFTPVNSLDPPLLTRYLRIHPOSWYHOIALRMYTGCEADOLY 2351
Db      2281 KVFQGNODSEFTPVNSLDPPLLTRYLRIHPOSWYHOIALRMYTGCEADOLY 2332

RESULT 14
PCT-US94-13200-2
; Sequence 2, Application PC/TUS9413200
; GENERAL INFORMATION:
; APPLICANT: Emory University
; TITLE OF INVENTION: Hybrid Human/Animal Factor VIII
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESS: Kilpatrick & Cody
; STREET: 1100 Peachtree Street, Suite 2800
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: US
; ZIP: 30309
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentia Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/13200
; FILING DATE: 15-NOV-1994
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Pabst, Patricia L.
; REGISTRATION NUMBER: 31,284
; REFERENCE/DOCKET NUMBER: EMU106CIP(2)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404-815-6508
; TELEFAX: 404-815-6555
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2332 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: YES
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Homo sapien
; TISSUE type: Liver cDNA sequence

```

PCT-US94-13200-2

Query Match 99.1%; Score 12301; DB 5; Length 2332;

Best Local Similarity 99.9%; Pred. No. 0;

Matches 2330; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 20 ATRRYTLAGAVELSPWYQMSDGLPVDARPPRPVRSFPPWTVYVYKTLFVEPTDHPFN 79
DB 1 ATRRYTLAGAVELSPWYQMSDGLPVDARPPRPVRSFPPWTVYVYKTLFVEPTDHPFN 60

QY 80 IAKRPPMMGLGPITQAEVYDYVITIKMASHYVSLHANGVYKMASGAGYDQISQ 139
DB 61 IAKRPPMMGLGPITQAEVYDYVITIKMASHYVSLHANGVYKMASGAGYDQISQ 120

QY 140 REKEDKVPFGSGSHYVWQVLKENGPMASDPCLTYSTLSHVLDLYKDLNGLIGALLVCR 199
DB 121 REKEDKVPFGSGSHYVWQVLKENGPMASDPCLTYSTLSHVLDLYKDLNGLIGALLVCR 180

QY 200 EGSIAKEKTOTLHKFTLLFAVFDGSKSMHSETRKNSLMODRDAASARAMPKHTVNGVYNR 259
DB 181 EGSIAKEKTOTLHKFTLLFAVFDGSKSMHSETRKNSLMODRDAASARAMPKHTVNGVYNR 240

QY 260 SLPGILGCHRRKSVYWHYVGMGTTPENHSTPLRGHTFLYRNRQASLEISPPPTLTAQTL 319
DB 241 SLPGILGCHRRKSVYWHYVGMGTTPENHSTPLRGHTFLYRNRQASLEISPPPTLTAQTL 300

QY 320 MDLGOFLFCHTISHQHBGMEAVYKVDSCPEEPOLRMKNNEADYDDDLTDEMDVYRF 379
DB 301 MDLGOFLFCHTISHQHBGMEAVYKVDSCPEEPOLRMKNNEADYDDDLTDEMDVYRF 360

QY 380 DDNSPSFIQIRSVAKKHPTWVHYIAEBEDMDYAPLYLAPDDRSTKSOYLKNGFORIG 439
DB 361 DDNSPSFIQIRSVAKKHPTWVHYIAEBEDMDYAPLYLAPDDRSTKSOYLKNGFORIG 420

QY 440 RYKKYRMAVYDDEFKREALIOHESGILGPIIYGEVGDITLITFKQASRPVNIYPHCI 499
DB 421 RYKKYRMAVYDDEFKREALIOHESGILGPIIYGEVGDITLITFKQASRPVNIYPHCI 480

QY 500 TVVRPLYSRRLPKGVKHLKDEPILIGEIFRYKMYTVVEDGPTKSDPRCLARTYSSFVNE 559
DB 481 TVVRPLYSRRLPKGVKHLKDEPILIGEIFRYKMYTVVEDGPTKSDPRCLARTYSSFVNE 540

QY 560 RDLASGLIGPLITCYKESVQDQGNQIMSDKRNVILFSVFDNRSMVLTENIQOFLNPNAG 619
DB 541 RDLASGLIGPLITCYKESVQDQGNQIMSDKRNVILFSVFDNRSMVLTENIQOFLNPNAG 600

QY 620 VOLEDPEFOASNIHMSINGYVFDLSVCLHRYAVYIISLGAQDPLSVFSSGTYFRH 679
DB 601 VOLEDPEFOASNIHMSINGYVFDLSVCLHRYAVYIISLGAQDPLSVFSSGTYFRH 660

QY 680 KMYEDTLLTFPFSGEVYFMSMENPGLMILGCHNSDFNRGMTALLVYSSCDKNTGDYFE 739
DB 661 KMYEDTLLTFPFSGEVYFMSMENPGLMILGCHNSDFNRGMTALLVYSSCDKNTGDYFE 720

QY 740 DSYEDISATLLSKNNAIEPRFSONSHPSTOKOFNATITTPENDIKETPPWAHHTPMP 799
DB 721 DSYEDISATLLSKNNAIEPRFSONSHPSTOKOFNATITTPENDIKETPPWAHHTPMP 780

QY 800 KIONVSSDLMLMLQOSTPHGSLSDLOBAKTYFSDPSGAIIDSNNLSLMTHERPQ 859
DB 781 KIONVSSDLMLMLQOSTPHGSLSDLOBAKTYFSDPSGAIIDSNNLSLMTHERPQ 840

QY 860 LHHSGDMVFTPESGILQELNKEKLGTTAAELKLDPKVYSTSNMLISTIPSDMLAAGTDN 919
DB 841 LHHSGDMVFTPESGILQELNKEKLGTTAAELKLDPKVYSTSNMLISTIPSDMLAAGTDN 900

QY 920 TSSLGPPSMVHYVDSQDITLFGKKSAPLLEGGSLSLSENNDSKLLSEGLMNSQESGM 979
DB 901 TSSLGPPSMVHYVDSQDITLFGKKSAPLLEGGSLSLSENNDSKLLSEGLMNSQESGM 960

QY 980 GKAVSSTESGRLFKGKRAHGPALITKDNALFVYSISLTKTKTSNNSATNRKTHIDGPSL 1039
DB 961 GKAVSSTESGRLFKGKRAHGPALITKDNALFVYSISLTKTKTSNNSATNRKTHIDGPSL 1020

QY 1040 LIENSFWONILLESDEFEKKVTPILIHDMKMDKNAGALRLNHMSKKTSSKNMEMVOOK 1099
DB 1021 LIENSFWONILLESDEFEKKVTPILIHDMKMDKNAGALRLNHMSKKTSSKNMEMVOOK 1080

QY 1100 KEGPIPPDAONPDMSPFRLFLPESARWIOPTHGKNSLNSGQSPKOLVSGPKPSYEG 1159
DB 1081 KEGPIPPDAONPDMSPFRLFLPESARWIOPTHGKNSLNSGQSPKOLVSGPKPSYEG 1140

QY 1160 QNPLSKKNVYVGGERTKQVGLKEVAFSSNNLEFLTMDLHNNTHNOEKTOEELEK 1219
DB 1141 QNPLSKKNVYVGGERTKQVGLKEVAFSSNNLEFLTMDLHNNTHNOEKTOEELEK 1200

QY 1220 KETLIQENVYLPQIHVYVGRKNMKMLFLSTRONVEGSDYGAAPVLDPFRSLDSTNR 1279
DB 1201 KETLIQENVYLPQIHVYVGRKNMKMLFLSTRONVEGSDYGAAPVLDPFRSLDSTNR 1260

QY 1280 TKKHTAHFSKGEENLEGLGNQTKQIYEKACTRISPTNSOONPVTQSRKALKOPRL 1339
DB 1261 TKKHTAHFSKGEENLEGLGNQTKQIYEKACTRISPTNSOONPVTQSRKALKOPRL 1320

QY 1340 PLEETLEKRIIYDDTSTONSKNNKHLPTSTLOTIDYNEKGAITIOSPLSCLTRSHSI 1399
DB 1321 PLEETLEKRIIYDDTSTONSKNNKHLPTSTLOTIDYNEKGAITIOSPLSCLTRSHSI 1380

QY 1400 POANRSPLEIAVSSFSIRPIYLTREYLEQDNSSHLPAASYRKKGDSGVOSHSFLOGAKK 1459
DB 1381 POANRSPLEIAVSSFSIRPIYLTREYLEQDNSSHLPAASYRKKGDSGVOSHSFLOGAKK 1440

QY 1460 NNLSTALITLLEMTGQDQREVSGISATNSVYTKKVENYVLPKPLPRTSGAKELLPKVHI 1519
DB 1441 NNLSTALITLLEMTGQDQREVSGISATNSVYTKKVENYVLPKPLPRTSGAKELLPKVHI 1500

QY 1520 YQKDLPRTTSNGSGHLDYBESLSLOGEGAIKKNENRGPRLPYATATSSAKTSSK 1579
DB 1501 YQKDLPRTTSNGSGHLDYBESLSLOGEGAIKKNENRGPRLPYATATSSAKTSSK 1560

QY 1580 LLDPLAMDNHGTQIPIKEBWSQEKSPERTAAKRRKKTLLSLINACSNHAAIAAINEGONKP 1639
DB 1561 LLDPLAMDNHGTQIPIKEBWSQEKSPERTAAKRRKKTLLSLINACSNHAAIAAINEGONKP 1620

QY 1640 EIEVTAQGRTERLCSQNPVLRKRORETTTTLQSQDEIDVDITISEVKKKEDPDY 1699
DB 1621 EIEVTAQGRTERLCSQNPVLRKRORETTTTLQSQDEIDVDITISEVKKKEDPDY 1680

QY 1700 DEDENQSPSSFOKTRHYFLAVERLMDYGSSSPHYLRNPAQSGSYPOFKRYVPOERTD 1759
DB 1681 DEDENQSPSSFOKTRHYFLAVERLMDYGSSSPHYLRNPAQSGSYPOFKRYVPOERTD 1740

QY 1760 GSPTOPLYRGELENHGLGPIYIRAEVEDNIWVTRNOASRPYSFSSLSIYEEDOROGA 1819
DB 1741 GSPTOPLYRGELENHGLGPIYIRAEVEDNIWVTRNOASRPYSFSSLSIYEEDOROGA 1800

QY 1820 EPRKNVYKNETKTYFKMKVOHHMAPIKDFPDCKAAYSDVDLEKYHSGGLIGPLLYCHT 1879
DB 1801 EPRKNVYKNETKTYFKMKVOHHMAPIKDFPDCKAAYSDVDLEKYHSGGLIGPLLYCHT 1860

QY 1880 NTLNPAHQVQVDEPALLFTITDETCKSWYFENNERNCAPACNIQMEDPTKENYRFA 1939
DB 1861 NTLNPAHQVQVDEPALLFTITDETCKSWYFENNERNCAPACNIQMEDPTKENYRFA 1920

QY 1940 INGYIMDTLPGLVNAQOORIRMYLILSMGSENIHSHIRSGHVFYRKKEBYKMALYNXP 1999
DB 1921 INGYIMDTLPGLVNAQOORIRMYLILSMGSENIHSHIRSGHVFYRKKEBYKMALYNXP 1980

QY 2000 GVEEYVEMIPSKAGIWEVECLIGEBHIAAGMSLFLYVNSCOMPILMAAGHTRDQIRAS 2059
DB 1981 GVEEYVEMIPSKAGIWEVECLIGEBHIAAGMSLFLYVNSCOMPILMAAGHTRDQIRAS 2040

QY 2060 GOYGWAPRLARLHYSGINAMSTKEPESWIKVDLLAMIIHGIKTQGAROKFSSLYTSQ 2119
DB 2041 GOYGWAPRLARLHYSGINAMSTKEPESWIKVDLLAMIIHGIKTQGAROKFSSLYTSQ 2100

QY	2120	FIWISLDGKKQWYVNGSTGLWVFGVNDSSGKHNENPDIATYIHLPTHTSIRS	2179
Db	2101	FIWISLDGKKQWYVNGSTGLWVFGVNDSSGKHNENPDIATYIHLPTHTSIRS	2160
QY	2180	TLRHELMGCDLNSCSMPHIGESKASIDAQITASSYFTNNFATWSPBARLHLQGRSNMR	2239
Db	2161	TLRHELMGCDLNSCSMPHIGESKASIDAQITASSYFTNNFATWSPBARLHLQGRSNMR	2220
QY	2240	POVNNPKEMLDQVFOFKTMKVTGVTTCGVKSLTSMTYKREFFLISSSGOHQWTLFPNGKV	2289
Db	2221	POVNNPKEMLDQVFOFKTMKVTGVTTCGVKSLTSMTYKREFFLISSSGOHQWTLFPNGKV	2280
QY	2300	KVFGQNDSTFPVNSLDPPLTLTKLRIHQSWHDIARNEYLGCEADLY	2351
Db	2281	KVFGQNDSTFPVNSLDPPLTLTKLRIHQSWHDIARNEYLGCEADLY	2332
RESULT 15			
US-08-276-594A-2			
: Sequence 2, Application US/08276594A			
: Patent No. 5693499			
: GENERAL INFORMATION:			
: APPLICANT: YONEMURA, Hitoshi			
: APPLICANT: TRAJIMA, Yoshitaka			
: APPLICANT: SUGAMARA, Keishin			
: APPLICANT: MASUDA, Kenichi			
: TITLE OF INVENTION: PROCESS FOR PREPARING HUMAN COAGULATION			
: TITLE OF INVENTION: FACTOR VIII PROTEIN COMPLEX			
: NUMBER OF SEQUENCES: 11			
: CORRESPONDENCE ADDRESS:			
: ADDRESSEE: Foley & Lardner			
: STREET: 3000 K Street, N.W., Suite 500			
: CITY: Washington			
: STATE: D.C.			
: COUNTRY: USA			
: ZIP: 20007-5109			
: COMPUTER READABLE FORM:			
: MEDIUM TYPE: floppy disk			
: COMPUTER: IBM PC compatible			
: OPERATING SYSTEM: PC-DOS/MS-DOS			
: SOFTWARE: PatentIn Release #1.0, Version #1.30			
: CURRENT APPLICATION DATA:			
: APPLICATION NUMBER: US/08/276,594A			
: FILING DATE: 18-JUL-1994			
: CLASSIFICATION: 435			
: PRIOR APPLICATION DATA:			
: APPLICATION NUMBER: US 07/950,191			
: FILING DATE: 24-SEP-1992			
: PRIOR APPLICATION DATA: JP 243262/1991			
: APPLICATION NUMBER: JP 243262/1991			
: FILING DATE: 24-SEP-1991			
: ATTORNEY/AGENT INFORMATION:			
: NAME: WEGNER, Harold C.			
: REGISTRATION NUMBER: 25,258			
: REFERENCE/DOCKET NUMBER: 74129/195/NOVA			
: TELECOMMUNICATION INFORMATION:			
: TELEPHONE: (202)672-5300			
: TELEFAX: (202)672-5399			
: TELEX: 904136			
: INFORMATION FOR SEQ ID NO: 2:			
: SEQUENCE CHARACTERISTICS:			
: LENGTH: 2332 amino acids			
: TYPE: amino acid			
: TOPOLOGY: linear			
: MOLECULE TYPE: protein			
US-08-276-594A-2			
Query Match 99.0%: Score 12294: DB 1: Length 2332:			
Best local Similarity 99.9%: P-Id: No. 0:			
Matches 2339: Conservative 1: Mismatches 2: Indels 0: Gaps 0:			
QY	20	ATRRYYGAVELSMQWSDLGEPLVDARPPRPVPSSEPPNFSVYKKRLTFVEFTDHLFN	79

Qy 1160 QNFLEKKNKVVNGKEFTKDVGLKENVPSSNLEFLTNLMDNHNHNTNOEKKIOEBIEK 1219
Db 1141 QNFLEKKNKVVNGKEFTKDVGLKENVPSSNLEFLTNLMDNHNHNTNOEKKIOEBIEK 1200
Qy 1220 KETLIGENVVLPQIHTYTGTRKNFMKNLFLSTROMVEGSGYGAAPYLODFSLINDSTNR 1279
Db 1201 KETLIGENVVLPQIHTYTGTRKNFMKNLFLSTROMVEGSGYGAAPYLODFSLINDSTNR 1260
Qy 1280 TKKHTAHFSKKEEENLEGLGNOTKOIYKXCTTRISPTSOONVTYQSKRALKQRL 1339
Db 1261 TKKHTAHFSKKEEENLEGLGNOTKOIYKXCTTRISPTSOONVTYQSKRALKQRL 1320
Qy 1340 PLEETELEKRIIVDTSTOWSKNNKHLTPSTLTQIDYNEKEKGAITQSPUSDCLTRSHSI 1399
Db 1321 PLEETELEKRIIVDTSTOWSKNNKHLTPSTLTQIDYNEKEKGAITQSPUSDCLTRSHSI 1380
Qy 1400 POANRSPPLIAVSSPFSIRPIYLTFLVLPQDNSSHLPAASYRKKGDSGVQESSHFLQAKK 1459
Db 1381 POANRSPPLIAVSSPFSIRPIYLTFLVLPQDNSSHLPAASYRKKGDSGVQESSHFLQAKK 1440
Qy 1460 NNLSTALITLEBTGDQREVSIGTSATNSYTKKYENTVLPKPDLPKTSKVELLPKYHI 1519
Db 1441 NNLSTALITLEBTGDQREVSIGTSATNSYTKKYENTVLPKPDLPKTSKVELLPKYHI 1500
Qy 1520 YOKDLFPTETSNKSPGHLDLVEGSLQGTGGAIKKNEANRPGKVPFLRVATESAKTPSK 1579
Db 1501 YOKDLFPTETSNKSPGHLDLVEGSLQGTGGAIKKNEANRPGKVPFLRVATESAKTPSK 1560
Qy 1580 LUDPLAMDNHGTQIPKEENKSOEKSPKTAKKDITLISNCESNHAIALINEGONKP 1639
Db 1561 LUDPLAMDNHGTQIPKEENKSOEKSPKTAKKDITLISNCESNHAIALINEGONKP 1620
Qy 1640 ELEVTAOKGTERICSONPPVLAKRQREITRTTLQSDQEEIDYDTISYEMKKEDFDIY 1699
Db 1621 ELEVTAOKGTERICSONPPVLAKRQREITRTTLQSDQEEIDYDTISYEMKKEDFDIY 1680
Qy 1700 DEDENOSPRSQKTRHYFLAVERLMDYGMSSSPHYLRNRAQSGVPOFKKVPFOEFTD 1759
Db 1681 DEDENOSPRSQKTRHYFLAVERLMDYGMSSSPHYLRNRAQSGVPOFKKVPFOEFTD 1740
Qy 1760 GSFTOPLYNGELNENLIGLGYIRAVEEDNIWTPRNOASRPYSLSIYEEDOROGA 1819
Db 1741 GSFTOPLYNGELNENLIGLGYIRAVEEDNIWTPRNOASRPYSLSIYEEDOROGA 1800
Qy 1820 EPRKNEFKNETKTYFKVOHMAPTKDEPDKANAYPSVDLEKDVHSGLIGPLVCHT 1879
Db 1801 EPRKNEFKNETKTYFKVOHMAPTKDEPDKANAYPSVDLEKDVHSGLIGPLVCHT 1860
Qy 1880 NTLNPAHGRVYVOEFALFTIFDETKSMYFTENNERNCRAPCNIQMEDPTREKENTRPHA 1939
Db 1861 NTLNPAHGRVYVOEFALFTIFDETKSMYFTENNERNCRAPCNIQMEDPTREKENTRPHA 1920
Qy 1940 INGYIMDTLPGIYMAOORIRWYLLSNGSENHSHSFGHVTVYRKKEEYKMALYNLYP 1999
Db 1921 INGYIMDTLPGIYMAOORIRWYLLSNGSENHSHSFGHVTVYRKKEEYKMALYNLYP 1980
Qy 2000 GYFETVEMLPKSAIGIWRVECLIGEHLHAGMSTLFLVSNKQOTPLGMASGHTRDFOITAS 2059
Db 1981 GYFETVEMLPKSAIGIWRVECLIGEHLHAGMSTLFLVSNKQOTPLGMASGHTRDFOITAS 2040
Qy 2060 GYGOWAPKLARLHSSSINASTKPEFSWIKVDLAPMIIHGKITOGAROKFSSLYTSQ 2119
Db 2041 GYGOWAPKLARLHSSSINASTKPEFSWIKVDLAPMIIHGKITOGAROKFSSLYTSQ 2100
Qy 2120 FIIMYSLDGKKQYIRKNSGTGLMVEFGVNDSSGIKNHFNPIIARYIRLHPTYSIRS 2179
Db 2101 FIIMYSLDGKKQYIRKNSGTGLMVEFGVNDSSGIKNHFNPIIARYIRLHPTYSIRS 2160
Qy 2180 TLRMELMGDLNCSMPLGMSKAIIDAQITASSYFTNMFTWSPSKARLHLOGRSNAWR 2239
Db 2161 TLRMELMGDLNCSMPLGMSKAIIDAQITASSYFTNMFTWSPSKARLHLOGRSNAWR 2220

Search completed: July 2, 2003, 09:47:30
Job time : 45 secs

Qy 2240 PQVNNPKEMLOVDFOKTMKVTGYTTOGVKSLTTSKYKEEFLISSODGHQWTLFPONGKV 2299
Db 2221 PQVNNPKEMLOVDFOKTMKVTGYTTOGVKSLTTSKYKEEFLISSODGHQWTLFPONGKV 2280
Qy 2300 KYPOGNDSFTPVVNSLDPLLRRLRIRHPOSWHQIALMNEYLGCEADOLY 2351
Db 2281 KYPOGNDSFTPVVNSLDPLLRRLRIRHPOSWHQIALMNEYLGCEADOLY 2332

QY 241 AASARAPKMTYVNGYVNRSLPGJICGCHKRYTAVTGMCTTPPVMSITFEGHTFLVNRH 300
DB 241 AASARAPKMTYVNGYVNRSLPGJICGCHKRYTAVTGMCTTPPVMSITFEGHTFLVNRH 300
QY 301 ROASLEISPIFLPAQTLMDLQOFLFCHISSHODMEAYVAVSCPEEPOLRKMKNE 360
DB 301 ROASLEISPIFLPAQTLMDLQOFLFCHISSHODMEAYVAVSCPEEPOLRKMKNE 360
QY 361 EADYDDDLTDEMOYVAFDDONSPEFIQIRSAKHKPTKVHTVIAEEDMDVAPVIA 420
DB 361 EADYDDDLTDEMOYVAFDDONSPEFIQIRSAKHKPTKVHTVIAEEDMDVAPVIA 420
QY 421 PDRSTKSOYLANGPORIGRKRYKVRPAVYDETEKTRREALIOHSGILGFLYGEVDTL 480
DB 421 PDRSTKSOYLANGPORIGRKRYKVRPAVYDETEKTRREALIOHSGILGFLYGEVDTL 480
QY 481 LIIFKNOASRPYNIYPHGIDVBPPLYSRRLPGVAKHLDPFLLPGEIFKYKWTVVEDGP 540
DB 481 LIIFKNOASRPYNIYPHGIDVBPPLYSRRLPGVAKHLDPFLLPGEIFKYKWTVVEDGP 540
QY 541 TKSDPRCLTRRYSSPVNMBDLASGLIPLLCYKESVDORNOIMDKRNVITLSVDE 600
DB 541 TKSDPRCLTRRYSSPVNMBDLASGLIPLLCYKESVDORNOIMDKRNVITLSVDE 600
QY 601 NRSWYLTENIQEFLPNPAGVQLEDPFOASNMHSINGYVDSIQLSVCLHEVAYWYILS 660
DB 601 NRSWYLTENIQEFLPNPAGVQLEDPFOASNMHSINGYVDSIQLSVCLHEVAYWYILS 660
QY 661 IGAQTDLSVFPSSGYTFKHKKVYEDTLTFPFGSETFVMSMENGMILICHSNDRNG 720
DB 661 IGAQTDLSVFPSSGYTFKHKKVYEDTLTFPFGSETFVMSMENGMILICHSNDRNG 720
QY 721 MTALLKVSCKDNKNGDYEDSYEDIASYLLSKNNALTEPRSSONSRRHROKOFNATTI 780
DB 721 MTALLKVSCKDNKNGDYEDSYEDIASYLLSKNNALTEPRSSONSRRHROKOFNATTI 780
QY 781 PENDIKTDPMFAHRTPMFKIQNVSSDDLMLLRSPFHGLSLSDIQEAKYETESDDPS 840
DB 781 PENDIKTDPMFAHRTPMFKIQNVSSDDLMLLRSPFHGLSLSDIQEAKYETESDDPS 840
QY 841 PGALDSNNSLSLMEHFRPOLHSGDMWTFPESGLQRLNEKLGTTAATELKIDFVVSST 900
DB 841 PGALDSNNSLSLMEHFRPOLHSGDMWTFPESGLQRLNEKLGTTAATELKIDFVVSST 900
QY 901 SNNLITIPSDMLAGDNTNSSLGPRSMHYHDSQDQTLTGKASSPLTSGSPILSSEE 960
DB 901 SNNLITIPSDMLAGDNTNSSLGPRSMHYHDSQDQTLTGKASSPLTSGSPILSSEE 960
QY 961 NDSKILLESGLMNSQESWGNVSTESGRLEFKGRAGPALLTKDNALFKVYSILKTN 1020
DB 961 NDSKILLESGLMNSQESWGNVSTESGRLEFKGRAGPALLTKDNALFKVYSILKTN 1020
QY 1021 KTSNNSATNRKTHIDGFLIENSPPVMONILLESDETEKKTPTLIHRMLMDNATLRL 1080
DB 1021 KTSNNSATNRKTHIDGFLIENSPPVMONILLESDETEKKTPTLIHRMLMDNATLRL 1080
QY 1081 NMSNKTTSKMKMEVQCKEGPPLPDQONPMSFFKMLFIPESARAIQCTHCKNLSNG 1140
DB 1081 NMSNKTTSKMKMEVQCKEGPPLPDQONPMSFFKMLFIPESARAIQCTHCKNLSNG 1140
QY 1141 QGSPKQVLSLGPESVYGQNFLSEKKKYVVGKGFEKDVGLKJMVPPSSRNLFITLNDN 1200
DB 1141 QGSPKQVLSLGPESVYGQNFLSEKKKYVVGKGFEKDVGLKJMVPPSSRNLFITLNDN 1200
QY 1201 LHENNTHNOEKKIOEIEIEKKETLLOENVYLPOIHTVGTCKPMKNLFLSTRNVGSGSD 1260
DB 1201 LHENNTHNOEKKIOEIEIEKKETLLOENVYLPOIHTVGTCKPMKNLFLSTRNVGSGSD 1260
QY 1261 GAYAPVLODFRSLNDSTNRTKKHTAHFSKGEENLEGLAQOTQOIVEKYACTRISLPT 1320
DB 1261 GAYAPVLODFRSLNDSTNRTKKHTAHFSKGEENLEGLAQOTQOIVEKYACTRISLPT 1320
QY 1321 SQONFVTOGRKALQOFLPLEETELEKRLIYDDSTQMSNMKHLPPSTLQIDVNEKE 1380

DB 1321 SQONFVTOGRKALQOFLPLEETELEKRLIYDDSTQMSNMKHLPPSTLQIDVNEKE 1380
QY 1381 KGATOSPILSCLTRSHSIPQANRSPPLIAKVSFPPSIRPYLRLVAFODNSHUPAASY 1440
DB 1381 KGATOSPILSCLTRSHSIPQANRSPPLIAKVSFPPSIRPYLRLVAFODNSHUPAASY 1440
QY 1441 RKKDSGVQESSHFIQAGKKNNSLALTLTEMGQDREVSLGTSANSTYKRYVNTVLP 1500
DB 1441 RKKDSGVQESSHFIQAGKKNNSLALTLTEMGQDREVSLGTSANSTYKRYVNTVLP 1500
QY 1501 KPDLPKTSKAYELLPKYHIYQKDLFPTETSNQSPGHDLVBGSLGTEGAIKMEANRP 1560
DB 1501 KPDLPKTSKAYELLPKYHIYQKDLFPTETSNQSPGHDLVBGSLGTEGAIKMEANRP 1560
QY 1561 GKVPFLRATSSAKTSPSKLLDPLADNHYGTQIPKEEMKSQEKSPKTAFAKKDITLST 1620
DB 1561 GKVPFLRATSSAKTSPSKLLDPLADNHYGTQIPKEEMKSQEKSPKTAFAKKDITLST 1620
QY 1621 NCESNHAIAINFGOMKPELEVTNAKQERTERLCSQNPPLARHOREITPTTLOSQDE 1680
DB 1621 NCESNHAIAINFGOMKPELEVTNAKQERTERLCSQNPPLARHOREITPTTLOSQDE 1680
QY 1681 IDYDITISYEMKKEDFDIYDEDENSPNSFOKTRHFFIAAVERLMDYGMSSPHYLRNR 1740
DB 1681 IDYDITISYEMKKEDFDIYDEDENSPNSFOKTRHFFIAAVERLMDYGMSSPHYLRNR 1740
QY 1741 AOGSVPOFKKVVQOEFPTDGSFTQPLRYGELNEHLGLGPYIAVEEDNIMVTFRNQSR 1800
DB 1741 AOGSVPOFKKVVQOEFPTDGSFTQPLRYGELNEHLGLGPYIAVEEDNIMVTFRNQSR 1800
QY 1801 PYSFYSLSIYSEEDORGAEPKRNFKPHEKTYKMYOHHMPTDEDEDCAMAYFSDV 1860
DB 1801 PYSFYSLSIYSEEDORGAEPKRNFKPHEKTYKMYOHHMPTDEDEDCAMAYFSDV 1860
QY 1861 DLEKDVHSLGILPLVCHINTLPNARGOVYVQEFALFTTIDETKSWYTEMERNCR 1920
DB 1861 DLEKDVHSLGILPLVCHINTLPNARGOVYVQEFALFTTIDETKSWYTEMERNCR 1920
QY 1921 PCNIQMEDPTEKENVRRHAINGYIMDTLGLVMAODORIRMYILSMGSMENHSHFSGH 1980
DB 1921 PCNIQMEDPTEKENVRRHAINGYIMDTLGLVMAODORIRMYILSMGSMENHSHFSGH 1980
QY 1981 VFTVVRKKEEYKALYNLYPGVFEFVEMLPSKAGIWRKECILGELHAGSTLFLVYSNK 2040
DB 1981 VFTVVRKKEEYKALYNLYPGVFEFVEMLPSKAGIWRKECILGELHAGSTLFLVYSNK 2040
QY 2041 QZPLGNASGHIRDFOQTASGOYQMAPKILARLHSGSINASTKEPFSWIKYDLAPMI 2100
DB 2041 QZPLGNASGHIRDFOQTASGOYQMAPKILARLHSGSINASTKEPFSWIKYDLAPMI 2100
QY 2101 HGIKTQGAROKFSSLYTSOFTIYSLDGKKMOTYRGNSGTGLIWPFGNDSSGIRHNIFN 2160
DB 2101 HGIKTQGAROKFSSLYTSOFTIYSLDGKKMOTYRGNSGTGLIWPFGNDSSGIRHNIFN 2160
QY 2161 PPLIANYIRLAPHTYSTRSTLRKELMGCDLNSCMPJLAMESKALISDAQTASSTYTNFA 2220
DB 2161 PPLIANYIRLAPHTYSTRSTLRKELMGCDLNSCMPJLAMESKALISDAQTASSTYTNFA 2220
QY 2221 TWSPSKARLHLQGRSNMARPQVNNPKEMLQVDFOKTKMYVGVTTQGVKSLLSMAYKEFL 2280
DB 2221 TWSPSKARLHLQGRSNMARPQVNNPKEMLQVDFOKTKMYVGVTTQGVKSLLSMAYKEFL 2280
QY 2281 ISSSODGOWMTLFPONKRYVPOGNDSTTPVYNSLPBLLTYRLIHPQSNVHOLALRM 2340
DB 2281 ISSSODGOWMTLFPONKRYVPOGNDSTTPVYNSLPBLLTYRLIHPQSNVHOLALRM 2340
QY 2341 EYLGEADOLY 2351
DB 2341 EYLGEADOLY 2351

RESULT 2


```
US-09-957-641-2
; Sequence 2, Application US/09957641
; Publication No. US20020182670A1
; GENERAL INFORMATION:
; APPLICANT: Emory University
; TITLE OF INVENTION: MODIFIED FACTOR VIII
; FILE REFERENCE: 75-00
; CURRENT APPLICATION NUMBER: US/09/957,641
; PRIORITY FILING DATE: 2001-09-15
; PRIOR APPLICATION NUMBER: US 60/234047
; PRIOR FILING DATE: 2000-09-19
; PRIOR APPLICATION NUMBER: US 60/236460
; PRIOR FILING DATE: 2000-09-29
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 2
; LENGTH: 2332
; TYPE: PRP
; ORGANISM: Homo sapiens
US-09-957-641-2

Query Match      99.1%; Score 12301; DB 9; Length 2332;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2330; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Oy      20  ATRRYTIGAVELSMQYKOSDGLPVDARPPRVKSPFNTSVYKKTLEVEFDHLEFN 79
      1  ATRRYTIGAVELSMQYKOSDGLPVDARPPRVKSPFNTSVYKKTLEVEFDHLEFN 60
Oy      80  IAKRPKMGGLGPIQAEYTDYVITLKNASHVYSLAHGVSYKASGATYDQTSQ 139
      61  IAKRPKMGGLGPIQAEYTDYVITLKNASHVYSLAHGVSYKASGATYDQTSQ 120
Oy      140  REKEDKYPGGSHYVWVYKENGPMASDPLCTYSLSHVLDYKDLNGLIGALLVCR 199
      121  REKEDKYPGGSHYVWVYKENGPMASDPLCTYSLSHVLDYKDLNGLIGALLVCR 180
Oy      200  EGSIAKEKOTLHKFLLFAVFDGCKSMHSETKNSIMODRDAASARAPCHRTVMGTVNR 259
      181  EGSIAKEKOTLHKFLLFAVFDGCKSMHSETKNSIMODRDAASARAPCHRTVMGTVNR 240
Oy      260  SLRPLGCHGRKSVYVHYVGMGTPEVHSTFLRGHTFLVNRHROASLEISITTLAQTL 319
      241  SLRPLGCHGRKSVYVHYVGMGTPEVHSTFLRGHTFLVNRHROASLEISITTLAQTL 300
Oy      320  MDLGFILFCGISSHOHDMGEAYVYKVDSCPEEPOLRMKNNEEADYDDDLTDEMDVAF 379
      301  MDLGFILFCGISSHOHDMGEAYVYKVDSCPEEPOLRMKNNEEADYDDDLTDEMDVAF 360
Oy      380  DDNSPSFTQIRSVAKKHPTWYHYIAEEDMDYAPLYLAPDDRSTKSOYLNGFORIG 439
      361  DDNSPSFTQIRSVAKKHPTWYHYIAEEDMDYAPLYLAPDDRSTKSOYLNGFORIG 420
Oy      440  RKTKVRFMAVYDETETKREAIQIESGILGPLLGEVSDTLITFKNAQSRPNITPHGI 499
      421  RKTKVRFMAVYDETETKREAIQIESGILGPLLGEVSDTLITFKNAQSRPNITPHGI 480
Oy      500  TDVRLPLSRRLPKGVKXHLKDPILPGELFKYKWTYVYEDGPTKSDPCLTRYSYSSVNM 559
      481  TDVRLPLSRRLPKGVKXHLKDPILPGELFKYKWTYVYEDGPTKSDPCLTRYSYSSVNM 540
Oy      560  RDLASGLIGPLLCYKESVDQNGQIMSDKRNVLLEVFEDENRSMYLTENIQIFLNPAG 619
      541  RDLASGLIGPLLCYKESVDQNGQIMSDKRNVLLEVFEDENRSMYLTENIQIFLNPAG 600
Oy      620  VOLEDEPOASNMHMSINGYVFDLSQVCLAEVAVWYLLISGAQOTFLSTFFSGYTFKH 679
      601  VOLEDEPOASNMHMSINGYVFDLSQVCLAEVAVWYLLISGAQOTFLSTFFSGYTFKH 660
Oy      680  KMYEDTLTLPFSGEYTFMNMENRGLMILGCHNSDFRNRGRTALAKVSSCDKNTDGYE 739
      661  KMYEDTLTLPFSGEYTFMNMENRGLMILGCHNSDFRNRGRTALAKVSSCDKNTDGYE 720
```

```
Oy      740  DSYEDISAVLTKNNALIPNSFSQNSRHPSTROKQFNATTTIPENDLEKTDUPFAHRTMP 799
      721  DSYEDISAVLTKNNALIPNSFSQNSRHPSTROKQFNATTTIPENDLEKTDUPFAHRTMP 780
Oy      800  KIQNVSSDLMLTLRQSPPTPGLSLSDLEAKYEFSDPSPGADISNNLSLSEMTFRQ 859
      781  KIQNVSSDLMLTLRQSPPTPGLSLSDLEAKYEFSDPSPGADISNNLSLSEMTFRQ 840
Oy      860  LHSNGMVTTPSSGQLRLNENKGTTRAAELKLDPKVYSSTNNLJSTIPSDMLAGDIN 919
      841  LHSNGMVTTPSSGQLRLNENKGTTRAAELKLDPKVYSSTNNLJSTIPSDMLAGDIN 900
Oy      920  TSSIGPPSPVHYDSQDPTTLFGKSSPLTESGGLSSEENDSKLLESGLNNQSSSM 979
      901  TSSIGPPSPVHYDSQDPTTLFGKSSPLTESGGLSSEENDSKLLESGLNNQSSSM 960
Oy      980  GKNVSTESGRLFKGRAGHPALLTDMNLFFVYSTSLTKTKTSNNSATNRKTHIDGSL 1039
      961  GKNVSTESGRLFKGRAGHPALLTDMNLFFVYSTSLTKTKTSNNSATNRKTHIDGSL 1020
Oy      1040  LIENSPVQNLIESDTEFKKVTPLJHDMAMDKNATALRLNHSNKTTSKKNMEVQOK 1099
      1021  LIENSPVQNLIESDTEFKKVTPLJHDMAMDKNATALRLNHSNKTTSKKNMEVQOK 1080
Oy      1100  KEGPIPPDQNDMSFFMLFLPESARWIDRTGKNSLNSGOGSPKQVLSLGPESVVG 1159
      1081  KEGPIPPDQNDMSFFMLFLPESARWIDRTGKNSLNSGOGSPKQVLSLGPESVVG 1140
Oy      1160  ONFLSKRNVVYGKEPFDKVDGLKEMVFPSSRKLPLTNLDNLHENRTHNDEKTOEITK 1219
      1141  ONFLSKRNVVYGKEPFDKVDGLKEMVFPSSRKLPLTNLDNLHENRTHNDEKTOEITK 1200
Oy      1220  KETLIQENVALPOLHVTGTNFKNLFLLSTRONVSGSYDCAVAPYLQDFSLANDSTNR 1279
      1201  KETLIQENVALPOLHVTGTNFKNLFLLSTRONVSGSYDCAVAPYLQDFSLANDSTNR 1260
Oy      1280  TKRHTAFSGKEEENLGLQNTQKQIYEKACTPTISPTSOQNFYQSRKALROFPL 1339
      1261  TKRHTAFSGKEEENLGLQNTQKQIYEKACTPTISPTSOQNFYQSRKALROFPL 1320
Oy      1340  PLEETELERKRIIVDTSTQNSKNKHLTPTSTLQIDYNEKEGALTOGSPISDCLTSHSI 1399
      1321  PLEETELERKRIIVDTSTQNSKNKHLTPTSTLQIDYNEKEGALTOGSPISDCLTSHSI 1380
Oy      1400  POANRSPPLIAVYSSPISIRIYLVRLFPQDNSSHLPASVKKKDSGOVSSHFLQAKK 1459
      1381  POANRSPPLIAVYSSPISIRIYLVRLFPQDNSSHLPASVKKKDSGOVSSHFLQAKK 1440
Oy      1460  MNLSLALITLMTGQORVYGLSATSNTYKRYVNTYLPKPDLPRTSGVELLRVHI 1519
      1441  MNLSLALITLMTGQORVYGLSATSNTYKRYVNTYLPKPDLPRTSGVELLRVHI 1500
Oy      1520  YOKDLPPTETSNQSGHDLVYEGSLQGTGALTKNNEANRPGKVPFLVATTESAATPSK 1579
      1501  YOKDLPPTETSNQSGHDLVYEGSLQGTGALTKNNEANRPGKVPFLVATTESAATPSK 1560
Oy      1580  LLDPLANDNHGTQILPKEMMSQKSEKAPFAKKOTIISLACSNHIALAINEQONRP 1639
      1561  LLDPLANDNHGTQILPKEMMSQKSEKAPFAKKOTIISLACSNHIALAINEQONRP 1620
Oy      1640  EIEVTAQKQRTBRLCSQNPVYLKRHOEITRTYLOSDEEIDYDITSYMKKEDFDY 1699
      1621  EIEVTAQKQRTBRLCSQNPVYLKRHOEITRTYLOSDEEIDYDITSYMKKEDFDY 1680
Oy      1700  DEDENQSPRSFOKTRHYFIAAVERLMDYMSSSPVLNRAQSGSVQOFKVVPOEFTD 1759
      1681  DEDENQSPRSFOKTRHYFIAAVERLMDYMSSSPVLNRAQSGSVQOFKVVPOEFTD 1740
Oy      1760  GSFQPLRBELENEHGLIGPIYAAVEDMDIMTFRPNOASRPYSYSSLISYEDDQOGA 1819
      1741  GSFQPLRBELENEHGLIGPIYAAVEDMDIMTFRPNOASRPYSYSSLISYEDDQOGA 1800
Oy      1820  EPRKNFVKPNETKTYFWVOHMAAPTRDEDCAMAFYSVDVLEKDVHSGILGPLLCHT 1879
```

```

1801 EPRKNVKNPKNETKTYFMVYOHMAPTDEEDDCAMAYFSDVLEDDVHSGILGLVCHH 1860
1880 NTLNPAHGRQVYVQEPALFTTIPDETSMYFTENMRNCRACQNIQMDPPFKENYFHA 1939
1861 NTLNPAHGRQVYVQEPALFTTIPDETSMYFTENMRNCRACQNIQMDPPFKENYFHA 1920
1940 INGYMDTLPGLVMAODORIRMYLLSMGSDNHNHSHFGSHVYTRKKEEYKALYMLP 1999
1921 INGYMDTLPGLVMAODORIRMYLLSMGSDNHNHSHFGSHVYTRKKEEYKALYMLP 1980
2000 GVEFTEVEMLEPSKAGIRVECLIGEHLHAGMSTFLVYSNKCOTPLGMAHGIHDFQITAS 2059
1981 GVEFTEVEMLEPSKAGIRVECLIGEHLHAGMSTFLVYSNKCOTPLGMAHGIHDFQITAS 2040
2060 GQYGOMAPKLARLHVSQSIINASTKEPSPATKYDLAPMTHIGIKTQCAROKFSSLYISQ 2119
2041 GQYGOMAPKLARLHVSQSIINASTKEPSPATKYDLAPMTHIGIKTQCAROKFSSLYISQ 2100
2120 FIIMYSIDGKKMQTYRGNSTGTLMWFEGNVDSSGIKHINFPPIIARYIRLPHYSIRS 2179
2101 FIIMYSIDGKKMQTYRGNSTGTLMWFEGNVDSSGIKHINFPPIIARYIRLPHYSIRS 2160
2180 TLRMELMGCDLNCSPPLGEMSKAISDAQITASSYTFNMFATSPSKARLHLOGSNAMR 2239
2161 TLRMELMGCDLNCSPPLGEMSKAISDAQITASSYTFNMFATSPSKARLHLOGSNAMR 2220
2240 PQVANNPKEMLOVDFOCTMKVTVGTVOGYKSLTSMYKEFLISSODGHQWTLFPONGV 2299
2221 PQVANNPKEMLOVDFOCTMKVTVGTVOGYKSLTSMYKEFLISSODGHQWTLFPONGV 2280
2300 KVFQGNDSFTPVVNSLDPILLTRRLRTHPSWVHOJALRMEVLGEADQDY 2351
2281 KVFQGNDSFTPVVNSLDPILLTRRLRTHPSWVHOJALRMEVLGEADQDY 2232

```

```

TELECOMMUNICATION INFORMATION:
TELEPHONE: 303/499-8080
TELEFAX: 303/499-8089
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2332 amino acids
TYPE: amino acid
STRANDEDNESS: single
MOLECULE TYPE: protein
POPOLOGY: <unknown>
HYPOTHETICAL: YES
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
TISSUE TYPE: Liver
SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-10-187-319-2

Query Match 99.1%; Score 12301; DB 9; Length 2332;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2330; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

20 ATRRRYILGAVELSMQYOSDLGELPYDAFPVRVKSPPFTSVYAKTLFEFTDHLFN 79
1 ATRRYILGAVELSMQYOSDLGELPYDAFPVRVKSPPFTSVYAKTLFEFTDHLFN 60
80 IAKRPWMGLIGPTIOAEYVDVYITLKNMASHVSLHANGVSTKASGADYDQTSQ 139
61 IAKRPWMGLIGPTIOAEYVDVYITLKNMASHVSLHANGVSTKASGADYDQTSQ 120
140 REKEDKVPGGSHYVQVLEKENGMAASPDLCTYSYLSHVLDYKDLNGLIGALLVCR 199
121 REKEDKVPGGSHYVQVLEKENGMAASPDLCTYSYLSHVLDYKDLNGLIGALLVCR 180
200 EGSIAKEKQTLHKTILFAVDEGSMSEKNSLMODROASARAMPKMHVNGYVNR 259
181 EGSIAKEKQTLHKTILFAVDEGSMSEKNSLMODROASARAMPKMHVNGYVNR 240
260 SLPLGLICHRKSVYVHVGMTPEVHSIFLBCGHFELVNNHROASLEISPTFLVQTL 319
241 SLPLGLICHRKSVYVHVGMTPEVHSIFLBCGHFELVNNHROASLEISPTFLVQTL 300
320 MDIGOFILRCHTSHOHGMEAYVYVQSCPEEPOLRMKNNEAEVDYDGLTDEMDVYF 379
301 MDIGOFILRCHTSHOHGMEAYVYVQSCPEEPOLRMKNNEAEVDYDGLTDEMDVYF 360
380 DDNSPSEFQIRSVAKKHPTKWVHYIAEEEDMDYAPLVLAADDKSYKSOYLNNQDRI 439
361 DDNSPSEFQIRSVAKKHPTKWVHYIAEEEDMDYAPLVLAADDKSYKSOYLNNQDRI 420
440 RYKRYKRFMAVYDEFTKREAIQHEGSIPLLYGVGDTLLIIFKNQASPVNIYPHGI 499
421 RYKRYKRFMAVYDEFTKREAIQHEGSIPLLYGVGDTLLIIFKNQASPVNIYPHGI 480
500 TDVRLYSLRLEKGYKHLKDPILIPGEIFKKTATVVEGPKSPKPLRTYYSYFVNE 559
481 TDVRLYSLRLEKGYKHLKDPILIPGEIFKKTATVVEGPKSPKPLRTYYSYFVNE 540
560 RDLASGLIGPLILICKEYSVDQGNQMSDRNRYILFVSDENRSWYLTENIORLPNAG 619
541 RDLASGLIGPLILICKEYSVDQGNQMSDRNRYILFVSDENRSWYLTENIORLPNAG 600
620 VOLDEPFOASNMHSINGYVFDLSQVCLHVAWYLLISGATDPLSVFSGYTFKH 679
601 VOLDEPFOASNMHSINGYVFDLSQVCLHVAWYLLISGATDPLSVFSGYTFKH 660
680 KMYEDTLTFPESGTEVMSKENVGLMIIGCHNSDFRRNGWTALLKYSCKNTGDYE 739
661 KMYEDTLTFPESGTEVMSKENVGLMIIGCHNSDFRRNGWTALLKYSCKNTGDYE 720
740 DSYEDISAVYLSKNNALIEPRSSQNSRHPSTROKONATTIENDIETDWFHRTMP 799

```

```

Db      721 DYEEDISAIVLKSNNNAIEPRSEFSONSNRHPSTROKQOFNAATTIPENDIEKTDPMFAHRTMPM 780
Qy      800 KIONVSSDULMLLRLQSPTRPHGLSLSDLOEAKYEFSDDPSPGALDSNNLSLEMTHPRPQ 859
Db      781 KIONVSSDULMLLRLQSPTRPHGLSLSDLOEAKYEFSDDPSPGALDSNNLSLEMTHPRPQ 840
Qy      860 LHHSGMVFPPESGLOLRLNEKLTCTTAATLAKLADRVSTSNNTLSTIPSDNTLAAGTDN 919
Db      841 LHHSGMVFPPESGLOLRLNEKLTCTTAATLAKLADRVSTSNNTLSTIPSDNTLAAGTDN 900
Qy      920 TSSIGPPSMRVHYDSQDLDYTLFGKSSPLTESGPILESSENNDSKLTESGLMNSQSSW 979
Db      901 TSSIGPPSMRVHYDSQDLDYTLFGKSSPLTESGPILESSENNDSKLTESGLMNSQSSW 960
Qy      980 GKNVSTESGRLFKGRRAHGPALLTKNALFKVSI SLTNTKTSNNSXNKKHTIDGSL 1039
Db      961 GKNVSTESGRLFKGRRAHGPALLTKNALFKVSI SLTNTKTSNNSXNKKHTIDGSL 1020
Qy      1040 LIENSPVQNTLESDETEFKKVTPLIHDRMLMDKNATALLRHNMSNKTSSKNMEVQK 1099
Db      1021 LIENSPVQNTLESDETEFKKVTPLIHDRMLMDKNATALLRHNMSNKTSSKNMEVQK 1080
Qy      1100 KEGPIPPDAQNPDMSFFKMLFLPESARWIOPTHGKSLNSGOGSPKOLVSLGPKSVEG 1159
Db      1081 KEGPIPPDAQNPDMSFFKMLFLPESARWIOPTHGKSLNSGOGSPKOLVSLGPKSVEG 1140
Qy      1160 QNFLSKKNVYVVGKEFTKDVGLKEWVPSSRNLFVNLDMJHENNTHNOEKKIOEBIEK 1219
Db      1141 QNFLSKKNVYVVGKEFTKDVGLKEWVPSSRNLFVNLDMJHENNTHNOEKKIOEBIEK 1200
Qy      1220 KETLIOENNVLPQIHTVYVGTKNFKNLFLSTNOVSGSYDGAIVPYLQDRSLNDSTNR 1279
Db      1201 KETLIOENNVLPQIHTVYVGTKNFKNLFLSTNOVSGSYDGAIVPYLQDRSLNDSTNR 1260
Qy      1280 TKKTHAHSKKEEENLEGAGNOTKQIYEKACTRISBNTSQONFYTORSKRALQOFRL 1339
Db      1261 TKKTHAHSKKEEENLEGAGNOTKQIYEKACTRISBNTSQONFYTORSKRALQOFRL 1320
Qy      1340 PLEETELEKRIIIVDTSTOMSKNMKHLTPSTLQIDYNEKKEGALTOPSIDCLTSHSI 1399
Db      1321 PLEETELEKRIIIVDTSTOMSKNMKHLTPSTLQIDYNEKKEGALTOPSIDCLTSHSI 1380
Qy      1400 POANSPPLPIAKVSSFPSIRPIYLTLYLQDNSSHLPAASTRKXDSGOVSSHFLOGARK 1459
Db      1381 POANSPPLPIAKVSSFPSIRPIYLTLYLQDNSSHLPAASTRKXDSGOVSSHFLOGARK 1440
Qy      1460 NNLSTAILTLEMTGDOREVGSLGTSATNSVTYKKTENTVLPKRPDLPTSGKVELLPKVIH 1519
Db      1441 NNLSTAILTLEMTGDOREVGSLGTSATNSVTYKKTENTVLPKRPDLPTSGKVELLPKVIH 1500
Qy      1520 YOKDLPEETNSGSPGHLDLVESSLQGTGEGAIKKNENARPCKVPLRVATESSANTPSK 1579
Db      1501 YOKDLPEETNSGSPGHLDLVESSLQGTGEGAIKKNENARPCKVPLRVATESSANTPSK 1560
Qy      1580 LLDPLAMDNHYGTOIPEEWRKSOEKSPERTAFKKDITLISLACESNHAIAAINEGONKP 1639
Db      1561 LLDPLAMDNHYGTOIPEEWRKSOEKSPERTAFKKDITLISLACESNHAIAAINEGONKP 1620
Qy      1640 EIEVTWAKOGTERELCSONPVYLKRHOREITRTLOSODEIDVDPTISVEMKKEPDY 1699
Db      1621 EIEVTWAKOGTERELCSONPVYLKRHOREITRTLOSODEIDVDPTISVEMKKEPDY 1680
Qy      1700 DEDENQSPRSFOKTRHYFTAAVERLMDYGMSSPHVLNRNAGSVPOPKKVVPOEFTD 1759
Db      1681 DEDENQSPRSFOKTRHYFTAAVERLMDYGMSSPHVLNRNAGSVPOPKKVVPOEFTD 1740
Qy      1760 GSFTOPLRGEINHEHLGLPYTRAVEEDNIWVTRNQAAPRYFSLSLSTYEDQROGA 1819
Db      1741 GSFTOPLRGEINHEHLGLPYTRAVEEDNIWVTRNQAAPRYFSLSLSTYEDQROGA 1800
Qy      1820 EPRKNFVAPNETKTYFKKVOHMAPTKDEFDCKAMAYFSVDLEKDVHSLIGPLLYCHT 1879
Db      1801 EPRKNFVAPNETKTYFKKVOHMAPTKDEFDCKAMAYFSVDLEKDVHSLIGPLLYCHT 1860

```

```

Qy      1880 NTLNPAHGRQYVOEPALEFTIEDERKSWTEENNERNCRAPCINOEDPTFEKENTRFA 1939
Db      1861 NTLNPAHGRQYVOEPALEFTIEDERKSWTEENNERNCRAPCINOEDPTFEKENTRFA 1920
Qy      1940 INGYIMDTLPGLVNAODORIRMYLILSMGSENENIHSIHFSGHVFYVRKEEYKNALYNLYP 1999
Db      1921 INGYIMDTLPGLVNAODORIRMYLILSMGSENENIHSIHFSGHVFYVRKEEYKNALYNLYP 1980
Qy      2000 GYFETVEMLPKACIMVECLIEHHAQMSLFLVYSNKCQTLVGAASHINDFOITAS 2059
Db      1981 GYFETVEMLPKACIMVECLIEHHAQMSLFLVYSNKCQTLVGAASHINDFOITAS 2040
Qy      2060 GQYGMARLARLHYSGSINAMSTKEPFSMIRVYDLAPMIIHGIKTQAGAROFSSLYIQ 2119
Db      2041 GQYGMARLARLHYSGSINAMSTKEPFSMIRVYDLAPMIIHGIKTQAGAROFSSLYIQ 2100
Qy      2120 FIIMTSLDGKKMOTYRONSOTGLMVFEGVDSGKIKHIFNPILARYILMHTHSIS 2179
Db      2101 FIIMTSLDGKKMOTYRONSOTGLMVFEGVDSGKIKHIFNPILARYILMHTHSIS 2160
Qy      2180 TLRMELMCCDINSCLMPLGMSKRAISDAQITASVFTNMFATMSPSKARLHLQGRSNMR 2239
Db      2161 TLRMELMCCDINSCLMPLGMSKRAISDAQITASVFTNMFATMSPSKARLHLQGRSNMR 2220
Qy      2240 POVNNPKEMLOVDFOKTMKVTYGTOTGVASLITSMYVEFLISSODGHOMTLFONGRY 2299
Db      2221 POVNNPKEMLOVDFOKTMKVTYGTOTGVASLITSMYVEFLISSODGHOMTLFONGRY 2280
Qy      2300 KYFOGNODSFTPVVNSLDPLLTRYLRIRHPOSWQOIALRMEVLCCEADOLY 2351
Db      2281 KYFOGNODSFTPVVNSLDPLLTRYLRIRHPOSWQOIALRMEVLCCEADOLY 2332

RESULT 4
US-10-187-319-6
: GENERAL INFORMATION:
: APPLICANT: Lollier, John S.
: TITLE OF INVENTION: Hybrid Human/Animal Factor VIII
: NUMBER OF SEQUENCES: 40
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Greenlee, Winner and Sullivan, P.C.
: STREET: 5370 Manhattan Circle Suite 201
: CITY: Boulder
: STATE: Colorado
: COUNTRY: USA
: ZIP: 80303
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/10/187, 319
: FILING DATE: 27-Aug-2002
: CLASSIFICATION: <Unknown>
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 09/533, 656
: FILING DATE: 2000-03-10
: APPLICATION NUMBER: US 09/037, 601
: FILING DATE: 1998-03-10
: APPLICATION NUMBER: WO PCT/US97/11155
: FILING DATE: 1997-06-26
: APPLICATION NUMBER: US 08/670, 707
: FILING DATE: 1996-06-26
: ATTORNEY/AGENT INFORMATION:
: NAME: Greenlee, Lorraine L.
: REGISTRATION NUMBER: 27, 894
: REFERENCE/DOCKET NUMBER: 75-95K
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 303/499-8080
: TELEFAX: 303/499-8089
: TITLE: Sequence of the Murine Factor VIII cDNA

```


Db 1938 ENIQSIHFSGHVTVKKEEYKMAVYNLYPGVFETLEMIPSRAGIWRVECLIGENHQAQM 1997
 QY 2030 STLFVYSNKCOTPLGMAISGHIIRDFOITASGOYQGMAPRLAHYSGSINAMSTKEPFM 2089
 Db 1998 STLFVYSKQCOIPLGMAISGSRDFOITASGHIQGMAPRLAHYSGSINAMSTKEPFM 2057
 QY 2090 IKYDLAPMHTIGITOCAROKESLISOFITIMYSLOKMKQYRGNISGTLMTFFGV 2149
 Db 2058 IKYDLAPMHTIGITOCAROKESLISOFITIMYSLOKMKQYRGNISGTLMTFFGV 2117
 QY 2150 DSGIKHNIFNPITARIYRLPHYSIRSLTLMELMGCDLNSCSMPLGMSKRAISDAOI 2209
 Db 2118 DSGIKHNIFNPITARIYRLPHYSIRSLTLMELMGCDLNSCSMPLGMSKRAISDAOI 2177
 QY 2210 TASGYTNPATWSPSKARLHOGSRNAMPQVNNPKRELVQVPOKTKATYGTQGVKS 2269
 Db 2178 TASGYTNPATWSPSKARLHOGSRNAMPQVNNPKRELVQVPOKTKATYGTQGVKS 2237
 QY 2270 LITSNYKEPILISSODGHQWTFPONGKATYQGNDSFTYVNSLDPPLTRYLRIRP 2329
 Db 2238 LITSNYKEPILISSODGHQWTFPONGKATYQGNDSFTYVNSLDPPLTRYLRIRP 2297
 QY 2330 QSWVHOIALMEYLGCDAODLY 2351
 Db 2298 QIWEHOIALREILGCEAOQY 2319
 RESULT 5
 US-10-187-319-37
 ; Sequence 37, Application US/10187319
 ; Publication No. US20030068785A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Lollar, John S.
 ; TITLE OF INVENTION: Hybrid Human/Animal Factor VIII
 ; NUMBER OF SEQUENCES: 40
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Greenlee, Winner and Sullivan, P.C.
 ; STREET: 5370 Manhattan Circle Suite 201
 ; City: Boulder
 ; STATE: Colorado
 ; COUNTRY: USA
 ; ZIP: 80303
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent in Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/10/187,319
 ; FILING DATE: 27-Aug-2002
 ; CLASSIFICATION: <unknown>
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 09/523,656
 ; FILING DATE: 2000-03-10
 ; APPLICATION NUMBER: US 09/037,601
 ; FILING DATE: 1998-03-10
 ; APPLICATION NUMBER: WO PCT/US97/11155
 ; FILING DATE: 1997-06-26
 ; APPLICATION NUMBER: US 08/670,707
 ; FILING DATE: 1996-06-26
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Greenlee, Lorance L.
 ; REGISTRATION NUMBER: 27,894
 ; REFERENCE/DOCKET NUMBER: 75-95K
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 303/499-8080
 ; TELEFAX: 303/499-8089
 ; INFORMATION FOR SEQ ID NO: 37:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2133 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein

; SEQUENCE DESCRIPTION: SEQ ID NO: 37:
 US-10-187-319-37
 Query Match 65.0%; Score 8076; DB 9; Length 2133;
 Best local similarity 67.0%; Pred. No. 0;
 Matches 1578; Conservative 208; Mismatches 343; Indels 226; Gaps 17;
 QY 1 MQTLSTCFELCLLRCSATRRYTLGAVELSDYQSD-LGEVYDARPPRPVPSFPE 59
 Db 1 MQTLSTCFELCLLRCSATRRYTLGAVELSDYQSD-LGEVYDARPPRPVPSFPE 60
 QY 60 NTSVYVKKTLFEEDTHLENIAKRPMPMGLPTIOAEVYTVVITLKNAASHVSLAH 119
 Db 61 GSVLYKKYVFEFTDQLESVARRPPMGLPTIOAEVYTVVITLKNAASHVSLAH 120
 QY 120 VGSVYVKSAGAEYDQTSQREKEDDYFPGGSHYVQVQVNLKENGPMASDPLCTYSYLS 179
 Db 121 VGSVYVKSAGAEYDQTSQREKEDDYFPGGSHYVQVQVNLKENGPMASDPLCTYSYLS 180
 QY 180 HVDLVKDLNSGLIGALLVCBSGLAKEKTQTLKFTLFAVDECKSHSETKNSLMQDR 239
 Db 181 HVDLVKDLNSGLIGALLVCBSGLAKEKTQTLKFTLFAVDECKSHSETKNSLMQDR 240
 QY 240 DAASARAMPKMTVNGYVRSILPGLIGCHRSYVMYVIGKGTPEVHSIFLEGHTFLVRN 299
 Db 241 DPAPARAQPMKMTVNGYVRSILPGLIGCHRSYVMYVIGKGTPEVHSIFLEGHTFLVRN 300
 QY 300 HROSLFESITPEFPAQTLMDGQFLFCHISHOHDMEAAYVVDSCREPOLRKNN 359
 Db 301 HROSLFESITPEFPAQTLMDGQFLFCHISHOHDMEAAYVVDSCREPOLRKNN 360
 QY 360 EEAEDYDDDLTJSEMDVVRFDSDNSPSEIOLRSVANKHPTVWHYIAAEEEDMDYAVLV 419
 Db 361 EE-EDYDNLTYDMDVVRFDSDNSPSEIOLRSVANKHPTVWHYIAAEEEDMDYAVLV 419
 QY 420 APPDRYSQYLNNGPQRIGRKRYKRVMAVTEDETPREDAIOHSGSLIGLGEVGDY 479
 Db 420 SPDSRSYKSLXLSNGPQRIGRKRYKRVMAVTEDETPREDAIOHSGSLIGLGEVGDY 479
 QY 480 LLIIFKQASRPYNYPHGIDTVRPLYSRLPQVYKHLKDFLLGELFKKRYVYEDG 539
 Db 480 LLIIFKQASRPYNYPHGIDTVRPLYSRLPQVYKHLKDFLLGELFKKRYVYEDG 539
 QY 540 PTKSDPCLLTRYYSFVNERNDLASGLIGPLLCYKESYDQGNQIMSDKRNVLISFVD 599
 Db 540 PTKSDPCLLTRYYSFVNERNDLASGLIGPLLCYKESYDQGNQIMSDKRNVLISFVD 599
 QY 600 EHSNMYLTEMIORELPNPGVQLEDEPERQASNTMHSINGVYFDSIQSLVCHAEVAYTIL 659
 Db 600 EHSNMYLTEMIORELPNPGVQLEDEPERQASNTMHSINGVYFDSIQSLVCHAEVAYTIL 659
 QY 720 GMTALLKVSQCDRIDIDYDNTEDIPGLISGKNVLEPSPASQNSRPSASQPOFTIT 779
 Db 720 GMTALLKVSQCDRIDIDYDNTEDIPGLISGKNVLEPSPASQNSRPSASQPOFTIT 779
 QY 780 IPENDIKETPPFAHNTMPKTIQNSSSDLMLLRQSPPHGISLSDQAEKAYTESDDP 839
 Db 780 IPENDIKETPPFAHNTMPKTIQNSSSDLMLLRQSPPHGISLSDQAEKAYTESDDP 839
 QY 840 SPGALDSNNSLSMETHFRPOLHSGDMVFTFPEGSLQRLNEKILGTTAATELKLDFVSS 899
 Db 840 SPGALDSNNSLSMETHFRPOLHSGDMVFTFPEGSLQRLNEKILGTTAATELKLDFVSS 899
 QY 837 LPGAERNTAPSAARLREPETHSAERLVLPREP-----ELKIKSS 882
 Db 837 LPGAERNTAPSAARLREPETHSAERLVLPREP-----ELKIKSS 882
 QY 900 TSNMLIS--TIPSDNLAAGTNTSSIGPPSPVAYVSDQLDITLFGKSSPPLTEGGGSL 957
 Db 900 TSNMLIS--TIPSDNLAAGTNTSSIGPPSPVAYVSDQLDITLFGKSSPPLTEGGGSL 957
 QY 883 SSDLKTSPIIPSDTISATERTHSLGPPHQQVAFNSQALAYLAKNSNHTIGAGVPLGS 942
 Db 883 SSDLKTSPIIPSDTISATERTHSLGPPHQQVAFNSQALAYLAKNSNHTIGAGVPLGS 942
 QY 938 SEENNDSKLLSGILNQSSESGKGVSTESGRLFKGRARHAPALLITDNLAFVYSISLL 1017
 Db 938 SEENNDSKLLSGILNQSSESGKGVSTESGRLFKGRARHAPALLITDNLAFVYSISLL 1017

```

Db      943 TEEDH-----ESSLGENVSPVESDGIFFEKERAHGPASTLTKDVLKFNISLV 989
Qy      1018 KTNTSNNSATNRKTHIDGSPILLSIENSPYQNIIESDTEFFKVPTLIDRLADOKNATA 1077
      990 KTNKARYTLKTNKTHIIDAALLTENRASA-----TFMDKNATA 1028
Qy      1078 LRLNHSNKTSSKNMKNMVOQKKESGIPPDQANDPMSFFMLFLPBSANWIORTHGKNSL 1137
      1029 SGLNHSN-----WIGPKKNPL 1047
Qy      1138 NSGQSPKQVSLGPEKSEVQONFLSEKNVYVVGKEFTKDYGLKEMVFPSSRLPLTN 1197
      1048 SSERGPPELLTSSGGKGVKGSGQGRIVAAVEBELSKG---KEMLPNSLPLTLN 1104
Qy      1198 LDMHNHTNOKKIOELEKKEKLLIOENVVLOIHTVGTNPKMLNLTSTQANEG 1257
      1105 SADVOGNDTHSOGKSEMEREREKLVQEKVLPQVYTAGTNNFLNLFHSTEPSVEG 1164
Qy      1258 SYDGAAPVLODFSLNDSTNFKKTHAFSKGEBENLEGLOTKOIVERYACTRIS 1317
      1165 FDGSHAPVPODSDSLNDSAEARETHIAHFAIREAPLEAPGNRT----- 1210
Qy      1318 PNTSQONTYQSRKRALQOERLPLETELEKRIIYDOSTOMSKNNKHLPLSTLQIDYN 1377
      1211 -GPGRSAPVPRKQOSKQIRLPLEIKERGVVLAATSTRMS----- 1252
Qy      1378 EKEKATQSPSLDCLTRSHSIPQANRSPPLIAVSSFPISIRPLYLRLVLFQDNSSHLPA 1437
      1253 ----- 1252
Qy      1438 ASYRKKGVOESHFLOGAKKNNLSLALILEMTGDORREVSIGSATNSVYKKEVNT 1497
      1253 -----ESSPILOGAKKNNLSLPLTLEMAAGGKISALGSAAGPLASGLEKA 1301
Qy      1498 VLPKPDLPKTSQVYELLPKVHIYQKDLFFETSNQSPHLDYVGSLLQSTBAIKMNEA 1557
      1302 VLSAGLSEAGKAEPLKPVYHREDLLPQKTSVNSCHAGDGLQELFLOKTRGPVNLNV 1361
Qy      1558 NRPKVPLRVAATESAKTSPKSLDPLANDNHYGTQIKREKMSQEKSPKTAFFAKKDTI 1617
      1362 NRPG-----RTPSKLGP-----PMK-EMESLEKSPSRTALTKDIT 1398
Qy      1618 -LSLACSNHAIATIEGONKPEIYVAKOGRERLCSNPNVYLKRHRERTRTTLOS 1676
      1399 SLPLRHSNHSIAKKEGQAEQREAAWTKGGGRCAKRPVLRHRORDISLTFOP 1458
Qy      1677 DOEIDYDITSVEMKKEDEDIYDEENQSPRSFOKTRHYFIAAVERLMDYGMSSSPHY 1736
      1459 EEDKMDYDIFSTFKGEDFDIYGEDENDPRSFOKTRHYFIAAVERLMDYGMSSSPRA 1518
Qy      1737 LNRRAQSSVQFKKVVFOEFTDSTOPLXYGELNEHLGILGYTIAEVEDNIMTFRN 1796
      1519 LNRRAQSSVQFKKVVFOEFTDSTOPLXYGELNEHLGILGYTIAEVEDNIMTFRN 1578
Qy      1797 QASRPYSFYSLLISYEDDQGAEPKRNKFNKRNKTYFWVOHHMAPTKDEDFCAWAY 1856
      1579 QASRPYSFYSLLISYEDDQGAEPKRNKFNKRNKTYFWVOHHMAPTKDEDFCAWAY 1638
Qy      1857 FSPVDEKDVHSGILGPLVCHNTLNPAHGOVYVOEALFFETIPDETSMVFTEKMER 1916
      1639 FSPVDEKDVHSGILGPLVCHNTLNPAHGOVYVOEALFFETIPDETSMVFTEKMER 1698
Qy      1917 NCRAPCNIMEDPYFKENYRPHALNGYIMDTPLGLVMAODORIMYLLSGMSNENHSH 1976
      1699 NCRAPCNIMEDPYFKENYRPHALNGYIMDTPLGLVMAODORIMYLLSGMSNENHSH 1758
Qy      1977 FSGHVTYKKEEYKEMALYNLYPGVEFTVEMLPKSKAGLRVACLGLHLAGNSTLFLY 2036
      1759 FSGHVTYKKEEYKEMALYNLYPGVEFTVEMLPKSKAGLRVACLGLHLAGNSTLFLY 1818
Qy      2037 SNKQOTPLGASGHTRDFQITASQIGQAPKIALRLHSGSINAWSTKEPSPWIKYDLA 2096
      1819 SKECQAPLGMASGRTRDFQITASQIGQAPKIALRLHSGSINAWSTKEPSPWIKYDLA 1878

```

```

Qy      2097 PMIHGKTQGAOKFFSGLYISOFIIMYSLDCKKQYTRGNSGTTLWFFGNVDSGKIH 2156
      1879 PMIHGKTQGAOKFFSGLYISOFIIMYSLDCKKQYTRGNSGTTLWFFGNVDSGKIH 1938
Qy      2157 NIFNPPIARVRLPHHYSTRSLRMLMGCDLNSCNPMLGESKASISAOITASSYPT 2216
      1939 NIFNPPIARVRLPHHYSTRSLRMLMGCDLNSCNPMLGESKASISAOITASSYPT 1998
Qy      2217 NMFATWSPKARLHLOHNSNAMPQVNNKREMLQVDFQTKYKTYGVTGVSLSLMTY 2276
      1999 NMFATWSPKARLHLOHNSNAMPQVNNKREMLQVDFQTKYKTYGVTGVSLSLMTY 2058
Qy      2277 KEFLISSQDGHQWTLFPONGKVVVEGQNDSEFTPVNSLDPPLTYRLRHPQSWHOC 2336
      2059 KEFLISSQDGHQWTLFPONGKVVVEGQNDSEFTPVNSLDPPLTYRLRHPQSWHOC 2118
Qy      2337 ALRMEVLCGEADLY 2351
      2119 ALRMEVLCGEADLY 2133
Db

```

RESULT 6
US-10-095-718-2
Sequence 2, Application US/10095718
Patient No. US20020131956A1
GENERAL INFORMATION:
APPLICANT: Walsh, Christopher
APPLICANT: Chao, Hengjun
APPLICANT: Burstein, Haim
APPLICANT: Lynch, Carmel
APPLICANT: Stepan, Tony
APPLICANT: Munson, Keith
TITLE OF INVENTION: Adeno-Associated Virus Vectors Encoding Factor VIII and
FILE REFERENCE: 35052/204375
CURRENT FILING DATE: 2002-03-12
PRIOR APPLICATION NUMBER: 09/689,430
PRIOR FILING DATE: 2001-08-22
PRIOR APPLICATION NUMBER: 60/158,780
PRIOR FILING DATE: 1999-10-12
NUMBER OF SEQ ID NOS: 5
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 2
LENGTH: 1471
TYPE: PRT
ORGANISM: Homo sapiens B-domain deleted factor VIII
FEATURE:
OTHER INFORMATION: Homo sapiens BDD FVIII
US-10-095-718-2

Query Match 59.7%; Score 7412; DB 12; Length 1471;
Best Local Similarity 62.5%; Pred No 0;
Matches 1470; Conservative 0; Mismatches 1; Indels 880; Gaps 1;

```

Qy      1 MOELSTCFELCLLRCSATRRYTLGAVELSDWYQMSDGLGELPVDAKPPRVKSPFN 60
      1 MOELSTCFELCLLRCSATRRYTLGAVELSDWYQMSDGLGELPVDAKPPRVKSPFN 60
Db
Qy      61 TSYYKKTLEFVEFTDLFIAPKRPMMGLGPTIOAEVYDVITLKNMASHPSLHAY 120
      61 TSYYKKTLEFVEFTDLFIAPKRPMMGLGPTIOAEVYDVITLKNMASHPSLHAY 120
Db
Qy      121 GVSYWKASBEAEVDDTQSRKEDKVPFGSHRYVQVYLKENGPAASDPLCTLYSLH 180
      121 GVSYWKASBEAEVDDTQSRKEDKVPFGSHRYVQVYLKENGPAASDPLCTLYSLH 180
Db
Qy      181 VDLVKDNLGSLGALLVCRGSLAKETQTLHKFILLFAVDEGKSMHSETKNSLMQDRD 240
      181 VDLVKDNLGSLGALLVCRGSLAKETQTLHKFILLFAVDEGKSMHSETKNSLMQDRD 240
Db
Qy      241 AASARAMPKHTVNGVYNSLPGLIGCHRSKVYVHWYIGMGTTEVHSLIFLBSGHTLVANH 300

```

|||||
241 AASARAMPKMTVNGYVNRSLPGLICHRKSYVWVIGMTPEVHSIFEGTFLVNRH 300
QY 301 ROASEISPTFLPAOLHMDLOFLFCHISSHODGMEAYVAVDSCPEEOLRKKNNE 360
Db 301 ROASEISPTFLPAOLHMDLOFLFCHISSHODGMEAYVAVDSCPEEOLRKKNNE 360
QY 361 EADYDODLTOSEMDYVAFDDNSPSFIOISYAKKHPTVWYIAAEEDMDYAPLYA 420
Db 361 EADYDODLTOSEMDYVAFDDNSPSFIOISYAKKHPTVWYIAAEEDMDYAPLYA 420
QY 421 PDDRSYKQYLNNGFORIGRKYKKVPMAYTDEFKTRREALIOHSGILGFLYGEVDTL 480
Db 421 PDDRSYKQYLNNGFORIGRKYKKVPMAYTDEFKTRREALIOHSGILGFLYGEVDTL 480
QY 481 LIIFFKQASRPYNTYPHGITDVPILYSRRLPGVYHLDKDFILPGEIYFKKMTVYEOEP 540
Db 481 LIIFFKQASRPYNTYPHGITDVPILYSRRLPGVYHLDKDFILPGEIYFKKMTVYEOEP 540
QY 541 TKSDPCLTRYYSSFVNNERDLASGLIGPLLICYESVDORGNOIMSDKNVILFSEVDE 600
Db 541 TKSDPCLTRYYSSFVNNERDLASGLIGPLLICYESVDORGNOIMSDKNVILFSEVDE 600
QY 601 NRSWYLTENIOFLPNPAGVOLEDPEFOASNIMHSINGYVDSIQLSVCLHEVAYWYILS 660
Db 601 NRSWYLTENIOFLPNPAGVOLEDPEFOASNIMHSINGYVDSIQLSVCLHEVAYWYILS 660
QY 661 IGAOTDELVSFSGYTFKHKKVYEDLTLPPFSGEYFMSKENGILICGHSNDRNG 720
Db 661 IGAOTDELVSFSGYTFKHKKVYEDLTLPPFSGEYFMSKENGILICGHSNDRNG 720
QY 721 MVALKVSCKDNKNGDYEDSEYEDISAYLISKNNALIEPRFSQNSRHPSTROKQFNATT 780
Db 721 MVALKVSCKDNKNGDYEDSEYEDISAYLISKNNALIEPRFSQNSRHPSTROKQFNATT 780
QY 781 PENDIEKTPWFAHRTPMRKIQNVSSSDLLMLKOSPTPHGLISDLQEAKEYTFSDPS 840
Db 779 ----- 778
QY 841 PGALDSNLSLMTFHRPOLHSGDMYTPDESQLOLNEKLGTTATELKIDFVSVST 900
Db 779 ----- 778
QY 901 SNMLISTIPSDNLAAGTNDTSSLCPPSPMAYVHDSOLDTTLFGKKSPLTESGGPLSLEE 960
Db 779 ----- 778
QY 961 NNDKILLESGLANSQESMSGKNVSTESGRLFEGKRAAGPALTRKNALFKVSIISLKTN 1020
Db 779 ----- 778
QY 1021 KTSNNSATNRKTHIDGPSLLIENSFWONILSDTEFKKYTPLIHDRMLMDKNAATL 1080
Db 779 ----- 778
QY 1081 NMSNKTSSKNMEVQOKKEGPIPPDAONPDMSFFKMLFLPESARMIORTHGKNSLNG 1140
Db 779 ----- 778
QY 1141 QGSPKQIVSLGPEKSEVQONFLSEKKNVYGGEFTKDVGLKEVFPSSRNLFJNLDN 1200
Db 779 ----- 778
QY 1201 LHENNTHNOEKIOEIEKKETLQENVYVLPOLHYTGKKNMKNLPLLSTRQWEGSYD 1260
Db 779 ----- 778
QY 1261 GAYAPVLODFRSLNDSTNRKTHAHFSKGEENLEGLGQTOQIYEKYACTTRISPMW 1320
Db 779 ----- 778
QY 1321 SQQNFYQSRSKRALQOFRLPLEETELBKRIYVDSTQSKNMKHLPLPSLTQIDYNEKE 1380

Db 779 ----- 778
QY 1381 KGAIQSPSLDCLTRSHSIPQANRSPILIAKVSPPSIRPIYLRVLQDNSSHLPAASY 1440
Db 779 ----- 778
QY 1441 RKDQGVQSSHFLOGAKKNNLSLITLBTMGDQREVGSIGTSATVSYTKKVENTVLP 1500
Db 779 ----- 778
QY 1501 KPDLPYTSKVELLPKVHIOKDLFPTETSGNSPGHLDLVGSLQETEGAIKMEANRP 1560
Db 779 ----- 778
QY 1561 GKVPFLVATESSAKTPSKLDPPLANDNHGQIPIREMKQSEKSPKTAFFKKOTIISL 1620
Db 779 ----- 778
QY 1621 NACESNHAIAINEGONKPEIETWMAKQTERELCSQNPVLAKRHORETRTTLQSDOE 1680
Db 779 ----- 1674KHOREITRTTLQSDOE 800
QY 1681 IDYDDTISYEMARKEDFIYDEDENOQSPRSFOKTRHYFLAAVERLMDYGSSSPHYLRNR 1740
Db 801 IDYDDTISYEMARKEDFIYDEDENOQSPRSFOKTRHYFLAAVERLMDYGSSSPHYLRNR 860
QY 1741 AOSGSVPORKKVYPOEFTDSFTQPLRGLNENHIGLGPYIRAEVEDNIMTFERRQASR 1800
Db 861 AOSGSVPORKKVYPOEFTDSFTQPLRGLNENHIGLGPYIRAEVEDNIMTFERRQASR 920
QY 1801 PYSPFSLISYEEDOROGAEPKRNKVPNETKTYFMKVOHMAPTKDEFCAMAYFSDV 1860
Db 921 PYSPFSLISYEEDOROGAEPKRNKVPNETKTYFMKVOHMAPTKDEFCAMAYFSDV 980
QY 1861 DLEKDYHSLGILPLVCHTNTLPAHGRQVYVQEFALFETTFDEFKSMYTEMENENRCA 1920
Db 981 DLEKDYHSLGILPLVCHTNTLPAHGRQVYVQEFALFETTFDEFKSMYTEMENENRCA 1040
QY 1921 PCNIOEMDPFKENRFAINGYIMOTLPGVYAADORIRRYLLSMGNSNHSIHFSCH 1980
Db 1041 PCNIOEMDPFKENRFAINGYIMOTLPGVYAADORIRRYLLSMGNSNHSIHFSCH 1100
QY 1981 VFTVRRKEEKALYNLYPGVFETVEMLPSKAGIMRVECLIGEHLAAGSTFLVYSNKC 2040
Db 1101 VFTVRRKEEKALYNLYPGVFETVEMLPSKAGIMRVECLIGEHLAAGSTFLVYSNKC 1160
QY 2041 QTPGMASSHIRDFQITASGOYGOMAPRLARLHYSGSINASTKEPFSWIKVDLAPMII 2100
Db 1161 QTPGMASSHIRDFQITASGOYGOMAPRLARLHYSGSINASTKEPFSWIKVDLAPMII 1220
QY 2101 HGKIGTGAOKFESSLYISOPTIAMSILGKEMOTYRGNSGTILMPFENDSSGIRKNTFN 2160
Db 1221 HGKIGTGAOKFESSLYISOPTIAMSILGKEMOTYRGNSGTILMPFENDSSGIRKNTFN 1280
QY 2161 PPIIARYIRLPHYHSIRSLRMLMGCDLNSCMLPGESKASISDAQITASSYFTNMFA 2220
Db 1281 PPIIARYIRLPHYHSIRSLRMLMGCDLNSCMLPGESKASISDAQITASSYFTNMFA 1340
QY 2221 TWSBKAHLLOGRSNANRPVONNPKEMLOVDOKMKMYTGVTQGVKSLSTSMYKEFL 2280
Db 1341 TWSBKAHLLOGRSNANRPVONNPKEMLOVDOKMKMYTGVTQGVKSLSTSMYKEFL 1400
QY 2281 ISSSOGHOMTLFPONGYKVPONOSFPPVNSLDPRLTRILRIHPOSWQOIALRM 2340
Db 1401 ISSSOGHOMTLFPONGYKVPONOSFPPVNSLDPRLTRILRIHPOSWQOIALRM 1460
QY 2341 EVLGCEAODLY 2351
Db 1461 EVLGCEAODLY 1471


```
; Sequence 1, Application US/10225900
; Publication No. US2003007752A1
; GENERAL INFORMATION:
; APPLICANT: Cho, Myung-Sam
; APPLICANT: Chan, Sham-Yuen
; APPLICANT: Kelsey, William
; APPLICANT: Yee, Helena
; TITLE OF INVENTION: Expression System for Factor VIII
; FILE REFERENCE: MSB-7255
; CURRENT APPLICATION NUMBER: US/10/225,900
; PRIOR FILING DATE: 2002-08-22
; PRIOR APPLICATION NUMBER: US/09/209,916
; PRIORITY DATE: 1998-12-10
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 1438
; TYPE: PRN
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Derived from
; OTHER INFORMATION: human factor VIII sequence
; US-10-225-900-1

Query Match      58.2%; Score 7227; DB 9; Length 1438;
Best Local Similarity 61.6%; Pred. No. 0;
Matches 1437; Conservative 0; Mismatches 1; Indels 894; Gaps 1;

QY 20 ATRRYVIGAVELSDMOSDDELVDARPPRPVPSFPPNTSVYKKTLEFEFTDLFN 79
DB 1 ATRRTVGAVALSDMOSDDELVDARPPRPVPSFPPNTSVYKKTLEFEFTDLFN 60
QY 80 IAKPRPMMGLGPTIOAEVYDTVYITLKNASHPVSLHAGVSYKASEGAEYDQTSQ 139
DB 61 IAKPRPMMGLGPTIOAEVYDTVYITLKNASHPVSLHAGVSYKASEGAEYDQTSQ 120
QY 140 REKEDKVPFGSHYVMOYLKENGPMASDPLCTYSYLSHVLVDNLSGILGALVCR 199
DB 121 REKEDKVPFGSHYVMOYLKENGPMASDPLCTYSYLSHVLVDNLSGILGALVCR 180
QY 200 EGSIAKERTOTLHFTLLPAVDEKSWHSEKSNIMODDASAAAMPKMTVNGVYNR 259
DB 181 EGSIAKERTOTLHFTLLPAVDEKSWHSEKSNIMODDASAAAMPKMTVNGVYNR 240
QY 260 SLPGILGCHRSKSYVHVIGMGTTPVEVHSIFLEGHTFLVRNHRQASLEISPTFLAQTL 319
DB 241 SLPGILGCHRSKSYVHVIGMGTTPVEVHSIFLEGHTFLVRNHRQASLEISPTFLAQTL 300
QY 320 MDLGFLLFCHISSHODGMEAYVYKDSCEPEPOLRMKNNEAEADYDDDLTSEMDVYRF 379
DB 301 MDLGFLLFCHISSHODGMEAYVYKDSCEPEPOLRMKNNEAEADYDDDLTSEMDVYRF 360
QY 380 DDNSPSFIOIRSAVAKHNPVTWVYIAAEEDMDVAPLVAPDDRYSKQYLLNNGFORIG 439
DB 361 DDNSPSFIOIRSAVAKHNPVTWVYIAAEEDMDVAPLVAPDDRYSKQYLLNNGFORIG 420
QY 440 RYKRYVPMATYDEFTREAIQHSGLIGPLLYEVEGDLLIFKQOARPNYIPHI 499
DB 421 RYKRYVPMATYDEFTREAIQHSGLIGPLLYEVEGDLLIFKQOARPNYIPHI 480
QY 500 TVRPLYSRRLPKGVKHLKDFPILPEIFRYKMTVVEDGPTKSDPCLTRYSSVFNNE 559
DB 481 TVRPLYSRRLPKGVKHLKDFPILPEIFRYKMTVVEDGPTKSDPCLTRYSSVFNNE 540
QY 560 RDLASGLIPPLICYKSYVDORGNQMSDKRNVILFVFDENRSMYLTENIQFRLNPAG 619
DB 541 RDLASGLIPPLICYKSYVDORGNQMSDKRNVILFVFDENRSMYLTENIQFRLNPAG 600
QY 620 VOLEDEEQASINIMSGVYFDSLQSYCLHNAVYITLIGATQPLSVFFSGTFFH 679
DB 601 VOLEDEEQASINIMSGVYFDSLQSYCLHNAVYITLIGATQPLSVFFSGTFFH 660
QY 680 KMVEDTTLTFPFSGETVFMSENPGLMILGCHNSDFRNRGKTALTKVSSCDKNTGDYFE 739
```

```
DB 661 KMVEDTTLTFPFSGETVFMSENPGLMILGCHNSDFRNRGKTALTKVSSCDKNTGDYFE 720
QY 740 DSEYEDISAYLLSKNNALIEPRFSQNSRHPSTKOKOFNATYIPENDIEKTDPMFAARTBMP 799
DB 721 DSEYEDISAYLLSKNNALIEPRFSQNSRHPSTKOKOFNATYIPENDIEKTDPMFAARTBMP 742
QY 800 KIQNVSSDLMMLRQSPYTHGSLSDQAEAKYETFSDDPSGALDSNNSLSBMTNFRPQ 859
DB 743 KIQNVSSDLMMLRQSPYTHGSLSDQAEAKYETFSDDPSGALDSNNSLSBMTNFRPQ 742
QY 860 LHSQDVAFTPESGILOLRNEKLGTAATLKLDFKVSSTNNLSTIPSDNLAAGTDN 919
DB 743 LHSQDVAFTPESGILOLRNEKLGTAATLKLDFKVSSTNNLSTIPSDNLAAGTDN 742
QY 920 TSSLPSPMPHYDSQDITLFPKSSPITESGPILSSENNDSKLLBSGLMNSQESSM 979
DB 743 TSSLPSPMPHYDSQDITLFPKSSPITESGPILSSENNDSKLLBSGLMNSQESSM 742
QY 980 GKNSSTESGILFKRAHGPALLTRDNALFKVYSISLTKTKTSNNSATNRKTHIDGDSL 1039
DB 743 GKNSSTESGILFKRAHGPALLTRDNALFKVYSISLTKTKTSNNSATNRKTHIDGDSL 742
QY 1040 LIENSPWQNLIESDTEFKKYPTLIHDMMLDKNATLRLNHSNKTSTSKNMEVQOK 1099
DB 743 LIENSPWQNLIESDTEFKKYPTLIHDMMLDKNATLRLNHSNKTSTSKNMEVQOK 742
QY 1100 KEGPIPPDAQNDMSFFKMLFLPESARWIOPTHGKNSLNSGQSPKQVLSLCPKESVSG 1159
DB 743 KEGPIPPDAQNDMSFFKMLFLPESARWIOPTHGKNSLNSGQSPKQVLSLCPKESVSG 742
QY 1160 QNPLESKNVYVGKEFTKQVGLKEWFPSSRNLETLNIDLHENTHNOEKKIOEIEK 1219
DB 743 QNPLESKNVYVGKEFTKQVGLKEWFPSSRNLETLNIDLHENTHNOEKKIOEIEK 742
QY 1220 KETLIOENVVLPQHTYVGTNFKNKLFLSTRQNVESYDGAAPVLQDFBSLNDSTNR 1279
DB 743 KETLIOENVVLPQHTYVGTNFKNKLFLSTRQNVESYDGAAPVLQDFBSLNDSTNR 742
QY 1280 TKRHTAFSKGEBENLGLNQTQIVKACTTRISPTSQNFYVYQSKRALKQFRL 1339
DB 743 TKRHTAFSKGEBENLGLNQTQIVKACTTRISPTSQNFYVYQSKRALKQFRL 742
QY 1340 PLBETLEKRIIVDDTSTQWSKNNKHLPTSLNQIDYNEKEGATQSPSLDCLTRSHST 1399
DB 743 PLBETLEKRIIVDDTSTQWSKNNKHLPTSLNQIDYNEKEGATQSPSLDCLTRSHST 742
QY 1400 POANRSPDLIAVSSFPISIRPYLTRVLFDONSHLPAASYKKDSGVQESSHFLQAKK 1459
DB 743 POANRSPDLIAVSSFPISIRPYLTRVLFDONSHLPAASYKKDSGVQESSHFLQAKK 742
QY 1460 NNLIALITLLEMTGQREYSLGTSATNSVYKKEVENTVLPKPDLPKTSKVELLPKVHI 1519
DB 743 NNLIALITLLEMTGQREYSLGTSATNSVYKKEVENTVLPKPDLPKTSKVELLPKVHI 742
QY 1520 YQKDLPETISNGSPGHLDLVESLLOGBALIKMBANRPGKVPFLAVNTASSAKTPSK 1579
DB 743 YQKDLPETISNGSPGHLDLVESLLOGBALIKMBANRPGKVPFLAVNTASSAKTPSK 742
QY 1580 LLDPLANDNHGTQIPKEEMKSQEKSPKTAFAKKKDTILSLNACESNHAIAINEGONKP 1639
DB 743 LLDPLANDNHGTQIPKEEMKSQEKSPKTAFAKKKDTILSLNACESNHAIAINEGONKP 742
QY 1640 EIEVTAKQGRTERLCSONPVLKRROREITRTTLOSDQBEIDYDDTISVEKKEDPDY 1699
DB 743 EIEVTAKQGRTERLCSONPVLKRROREITRTTLOSDQBEIDYDDTISVEKKEDPDY 742
QY 1700 DEDENOSRFSQKTRHYEIAAVERLMDYGMSSSPHYLRNRAQSGSVQFKKVVFOEFTD 1759
DB 787 DEDENOSRFSQKTRHYEIAAVERLMDYGMSSSPHYLRNRAQSGSVQFKKVVFOEFTD 846
QY 1760 GSTQPLRGELEHGLGPTIRAEVENDJNTYFRQASRPSTYSLSIYEEDQOGA 1819
DB 1760 GSTQPLRGELEHGLGPTIRAEVENDJNTYFRQASRPSTYSLSIYEEDQOGA 1819
```

```

Db      847 GSEFTOPLYRGELNHLGLLPYIRAEVEDNIMWTFRNOASRPYSFYSSLSIYEEDROGA 906
Qy      1820 EPRKNEVAPNIEKTYEFKVOHNAPTKDEFDCKAMAYFSVDLEKDYHSLIGPLLYCHT 1879
Db      907 EPRKNEVAPNIEKTYEFKVOHNAPTKDEFDCKAMAYFSVDLEKDYHSLIGPLLYCHT 966
Qy      1880 NTLNPAHGRQYVQEFLLFTIFDEKSKYFTEENNERCRAPOCICOMEOPTEKREMYRHA 1939
Db      967 NTLNPAHGRQYVQEFLLFTIFDEKSKYFTEENNERCRAPOCICOMEOPTEKREMYRHA 1026
Qy      1940 INGYIMDTLPGLVMAOQORIRMYLLSMGSENINISIHFSGHVTFVRKKEEYKALYNLP 1999
Db      1027 INGYIMDTLPGLVMAOQORIRMYLLSMGSENINISIHFSGHVTFVRKKEEYKALYNLP 1086
Qy      2000 GVEFTEVEMLPKAGIWRVECLIGENHJAGMSTLFYVSNKQCPFLGMASGHIREDOITAS 2059
Db      1087 GVEFTEVEMLPKAGIWRVECLIGENHJAGMSTLFYVSNKQCPFLGMASGHIREDOITAS 1146
Qy      2060 GOYGOMAPKLALHYSGSINNAKTEPESMIVYDILAPMIIHGKTOGAROKFSSLYTSO 2119
Db      1147 GOYGOMAPKLALHYSGSINNAKTEPESMIVYDILAPMIIHGKTOGAROKFSSLYTSO 1206
Qy      2120 FIIMYSLDGKWKQYRCNSTGTLMVFPFGVNDSSGIKHNIEPPIIARYIRLAPHTYSIRS 2179
Db      1207 FIIMYSLDGKWKQYRCNSTGTLMVFPFGVNDSSGIKHNIEPPIIARYIRLAPHTYSIRS 1266
Qy      2180 TLRMELMGCDLNSGCMPLGMSKALSDAQITASSYFTNMFAWSPSKARLHOGRSNMR 2239
Db      1267 TLRMELMGCDLNSGCMPLGMSKALSDAQITASSYFTNMFAWSPSKARLHOGRSNMR 1326
Qy      2240 POYNNKEMVQVDFQKTKMYGTGTTQGVKSLTSMYKEFLISSQGHQWTLFEFNGKAY 2299
Db      1327 POYNNKEMVQVDFQKTKMYGTGTTQGVKSLTSMYKEFLISSQGHQWTLFEFNGKAY 1386
Qy      2300 KYFGNODSFPPVNSLDPPLLTLYLRIHPOSVMHQIALRMEVLGCEADOLY 2351
Db      1387 KYFGNODSFPPVNSLDPPLLTLYLRIHPOSVMHQIALRMEVLGCEADOLY 1438

RESULT 8
US-10-006-091-1
; Sequence 1, Application US/10006091
; Patient No. US20020102730AT
; GENERAL INFORMATION:
; APPLICANT: Cho, Myung-Sam
; APPLICANT: Chan, Sham-Yuen
; APPLICANT: Kelsey, William
; APPLICANT: Yee, Helena
; TITLE OF INVENTION: Expression System for Factor VIII
; FILE REFERENCE: MSB-7255.1
; CURRENT APPLICATION NUMBER: US/10/006,091
; CURRENT FILING DATE: 2001-12-06
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 1438
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Derived from
; OTHER INFORMATION: human factor VIII sequence
US-10-006-091-1

Query Match      58.2%; Score 7227; DB 12; Length 1438;
Best Local Similarity 61.6%; Pred. No. 0;
Matches 1437; Conservative 0; Mismatches 1; Indels 894; Gaps 1;

```

```

Db      61 IAKRPPMAGLIGPTIOAEYDVTVTILKNMAHPVSLAAGVSYWKAAGEYDDQTSQ 120
Qy      140 REKEDDKVPGGSHYVQVYLKENGPMASDPLCLTYSYLSHVLDYKDLNSGLIGALLVCR 199
Db      121 REKEDDKVPGGSHYVQVYLKENGPMASDPLCLTYSYLSHVLDYKDLNSGLIGALLVCR 180
Qy      200 EGSIAEKQOTLHKRILILFANPEGGSNHSEKNSLSMDORAAARAMPKMHVNGYVR 259
Db      181 EGSIAEKQOTLHKRILILFANPEGGSNHSEKNSLSMDORAAARAMPKMHVNGYVR 240
Qy      260 SLPLGICHRKSVYWHVIGMGTTPVHSIFLDSGHTPLVNRHROASLEISPTFLAQTLL 319
Db      241 SLPLGICHRKSVYWHVIGMGTTPVHSIFLDSGHTPLVNRHROASLEISPTFLAQTLL 300
Qy      320 MDGQFLCCHISSHQHDMGMAVYKVDSCPEEPOLRMKNNEADAYDDDLTJSEMDYVAF 379
Db      301 MDGQFLCCHISSHQHDMGMAVYKVDSCPEEPOLRMKNNEADAYDDDLTJSEMDYVAF 360
Qy      380 DDNSPSFQIRSVAKKHKPKTWVYIAAEEEDMDYAPLVIAPODRSYKSOYLNNQFORIG 439
Db      361 DDNSPSFQIRSVAKKHKPKTWVYIAAEEEDMDYAPLVIAPODRSYKSOYLNNQFORIG 420
Qy      440 RYKRVKRYPAVYDEFTKREALIOHESGILAPLYGDEVGDTLLIFKNQASRPYNIYPHGI 499
Db      421 RYKRVKRYPAVYDEFTKREALIOHESGILAPLYGDEVGDTLLIFKNQASRPYNIYPHGI 480
Qy      500 TDVRLPYSRRLPKGVKHLKDPILPGEIFYKWTYVVEDGPTKSPDCLIRYSSPVNNE 559
Db      481 TDVRLPYSRRLPKGVKHLKDPILPGEIFYKWTYVVEDGPTKSPDCLIRYSSPVNNE 540
Qy      560 RDLASGLIGPLLCYKESVDORGNQJMSDKRNVILFVSPDERKSYLTENIQRLPNPAG 619
Db      541 RDLASGLIGPLLCYKESVDORGNQJMSDKRNVILFVSPDERKSYLTENIQRLPNPAG 600
Qy      620 VOLEDEPQASIMHMSINGYVFDLSQVCLAEVAVYTLISGAOTDFLSVFPSCGYTFKH 679
Db      601 VOLEDEPQASIMHMSINGYVFDLSQVCLAEVAVYTLISGAOTDFLSVFPSCGYTFKH 660
Qy      680 KMYEDTILTFPESGETVPMENKPMGLIIGCHNSPFRRCGTALLKRVSSCDNTDGYE 739
Db      661 KMYEDTILTFPESGETVPMENKPMGLIIGCHNSPFRRCGTALLKRVSSCDNTDGYE 720
Qy      740 DSYEDISAYILSKNNAIEPRSFSONSHRSTROKQFNATTPENDIEKTOPWRAHRTPPM 799
Db      721 DSYEDISAYILSKNNAIEPRSFSONSHRSTROKQFNATTPENDIEKTOPWRAHRTPPM 742
Qy      800 KIQNVSSSDILMLNQSPTPHGLSLDLOEAKYEPSPDPSGAIDSNNSLSKWHFRPQ 859
Db      743 ----- 742
Qy      860 LKHSQGVNFTPESGLOLNLNKEKLGTTAATLKKLIDPKVSTSNMLISTIPSDNLAQTDN 919
Db      743 ----- 742
Qy      920 TSSLGPPMFBVHYDQDQDTLFGKSSPLTESGGPLSISENNDSKILSGLMNSQSSM 979
Db      743 ----- 742
Qy      980 GKNVSTESGRLFKGRAGPALLTKNALFKVISLTKNTKSNSATNKRTHIDOPSL 1039
Db      743 ----- 742
Qy      1040 LIENSPSVWONLESDETEKVTPLIHDRMLMDKNATLRLNHSNKTTSKKNENYQOK 1099
Db      743 ----- 742
Qy      1100 KEGPIPPDAONPDMSEFKMLFLPESAMRIORHGKNSLNSGOGSPKOLVSLGPEKSVEG 1159
Db      743 ----- 742
Qy      1160 QNPLSEKKNVYVKGGEFTKIDVGLKENVFPSSRLFLTNLDNHNHTNDKRIQOEIRK 1219
Db      743 ----- 742

```

QY 1220 KETLIOENVLPQIHVYTGKFNKNEFLSTRQNVESGYDAVAPLQDPRSLNDSTNR 1279
DB 743 ----- 742
QY 1280 TKKHTAFSKGEEENBGLGNOQIYEVACTRISPTSOQNFVORSKRALKOFR 1339
DB 743 ----- 742
QY 1340 PLEETELEKRIIVDSTQWSNMKHLTPSLQIDYNEKEGALTQPSLDCLTRSHSI 1399
DB 743 ----- 742
QY 1400 POANRSPPLIAKVSFSPTRPIYLRVLFDQNSHLPAASYRKDSGVQESSHFLQAKK 1459
DB 743 ----- 742
QY 1460 NNLSLAILTEMTGDQREVSLGTSATNSVYKKVENTVLPKBDLPKTSKVELLPVHI 1519
DB 743 ----- 742
QY 1520 YQKDLPTETSGSPGHLDVYEGSLQCTEGAIKKNANRPGKVPPLKAVNESSAKTPSK 1579
DB 743 ----- 742
QY 1580 LLDPLANDNHGTQIPKEEMKSOEKSPKTAFFKKKDTILSLNACSNHAIANEGONKP 1639
DB 743 ----- 742
QY 1640 ELEVTAQGRTERLCSQNPVPLKRRHREITRTTLOSDEIDYDPTISVEKKKEDPDIY 1699
DB 743 ----- 742
QY 1700 DEDENOSPRSFQKTRHYTIAAVERLMDYGMSSSPHYLNRKASGVPOPKVYVQETD 1759
DB 787 DEDENOSPRSFQKTRHYTIAAVERLMDYGMSSSPHYLNRKASGVPOPKVYVQETD 846
QY 1760 GSFQPLRGELNEHGLGPIYRAVEDNIMVFRNQSAPRSPFSSLSIESTEORGA 1819
DB 847 GSFQPLRGELNEHGLGPIYRAVEDNIMVFRNQSAPRSPFSSLSIESTEORGA 906
QY 1820 EPRKNFYKPNETKTYTWKVOHMAPTKDEPCKAMAFSVYDLEKYHSGLLGPIYVHT 1879
DB 907 EPRKNFYKPNETKTYTWKVOHMAPTKDEPCKAMAFSVYDLEKYHSGLLGPIYVHT 966
QY 1880 NTLNPAHROVTVQEFALFETTFDETKSMYFTENNERCAPCNIOMEDPTREKRYRHA 1939
DB 967 NTLNPAHROVTVQEFALFETTFDETKSMYFTENNERCAPCNIOMEDPTREKRYRHA 1026
QY 1940 INGTIMOTLPGIYVADODRIRWYLLSMGSENIHSHIFSGHFTVARKKEEYMALYNXP 1999
DB 1027 INGTIMOTLPGIYVADODRIRWYLLSMGSENIHSHIFSGHFTVARKKEEYMALYNXP 1086
QY 2000 GVFEVEMLPKAGIWEVECLIGEHLHAGMSTLELVYSNKCQTPPLGMAHGHTRDQITAS 2059
DB 1087 GVFEVEMLPKAGIWEVECLIGEHLHAGMSTLELVYSNKCQTPPLGMAHGHTRDQITAS 1146
QY 2060 GOYGMAPKLARLHYSSINAMSTKPEFSMIKVDLAPMIHIGIKQAGROKFSLSIYSQ 2119
DB 1147 GOYGMAPKLARLHYSSINAMSTKPEFSMIKVDLAPMIHIGIKQAGROKFSLSIYSQ 1206
QY 2120 FLIMTSLDGKMWQYRGNSTGTILMFGVNDSSGKININPPLIARYIRLHPHYSTRS 2179
DB 1207 FLIMTSLDGKMWQYRGNSTGTILMFGVNDSSGKININPPLIARYIRLHPHYSTRS 1266
QY 2180 TLRMELMGCDLNSGSMPLGMSKALISDAOITASSYFTNMFATSPSARLHLQGRSNMR 2239
DB 1267 TLRMELMGCDLNSGSMPLGMSKALISDAOITASSYFTNMFATSPSARLHLQGRSNMR 1326
QY 2240 PQVNNPKEMLOVDFOKTMKVTGVTGQVKSLLTSMYKREPLISSQDGHQWTLFQNGKV 2299
DB 1327 PQVNNPKEMLOVDFOKTMKVTGVTGQVKSLLTSMYKREPLISSQDGHQWTLFQNGKV 1386

QY 2300 KYFGQNDSTFPVYNSIDPPLRLRLRHPSWYHQAIALRMEYLGCEADOLY 2351
DB 1387 KYFGQNDSTFPVYNSIDPPLRLRLRHPSWYHQAIALRMEYLGCEADOLY 1438

RESULT 9
US-10-047-257-1
; Sequence 1, Application US/10047257
; Patent No. US2002011512A1
; GENERAL INFORMATION:
; APPLICANT: Cho, Myung-Sam
; APPLICANT: Chan, Shun-Yuen
; APPLICANT: Kelsey, William
; APPLICANT: Yee, Helena
; TITLE OF INVENTION: Expression System for Factor VIII
; FILE REFERENCE: MSB-725.2
; CURRENT FILING DATE: 2002-01-15
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 1438
; TYPE: PR
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Derived from
US-10-047-257-1
human factor VIII sequence

Query Match 58.2%; Score 7227; DB 12; Length 1438;
Best Local Similarity 61.6%; Pred. No. 0;
Matches 1437; Conservative 0; Mismatches 1; Indels 894; Gaps 1;

QY 20 ATRRYTLGAVELSDYKOSDGLPYDAPRPVPSKSPFNTSVYKTLFVETDHLFN 79
DB 1 ATRRYTLGAVELSDYKOSDGLPYDAPRPVPSKSPFNTSVYKTLFVETDHLFN 60
QY 80 IAKRPWNGLLGPTIAEYDVYVITLKNMASHPVSLHAGVSYKASGAEYDDQTSQ 139
DB 61 IAKRPWNGLLGPTIAEYDVYVITLKNMASHPVSLHAGVSYKASGAEYDDQTSQ 120
QY 140 REKEDKVPFGSHYVYQVLEKNGPASPCLTYSTYSHVDLYKDLNSGLGALLYCR 199
DB 121 REKEDKVPFGSHYVYQVLEKNGPASPCLTYSTYSHVDLYKDLNSGLGALLYCR 180
QY 200 EGSIAEKTQTLAKFTLLFAVDEGKSMSETKNSLMODRDAASARAPKMTYNGYVR 259
DB 181 EGSIAEKTQTLAKFTLLFAVDEGKSMSETKNSLMODRDAASARAPKMTYNGYVR 240
QY 260 SLPLGICHRKSVYVHVGMTPEVHSIFLEGHTFLVNRQASLEISPTLETLAQTL 319
DB 241 SLPLGICHRKSVYVHVGMTPEVHSIFLEGHTFLVNRQASLEISPTLETLAQTL 300
QY 320 MDLAGOTLCHSSHQHGMAAYKYVDSCEEPQLMKKNNEAEYDDDLDSMDVYR 379
DB 301 MDLAGOTLCHSSHQHGMAAYKYVDSCEEPQLMKKNNEAEYDDDLDSMDVYR 360
QY 380 DDNSPSFIOISVAKKHPKTMVYIAAEEEDVYAPLADBDKSYKQYLANGPORG 439
DB 361 DDNSPSFIOISVAKKHPKTMVYIAAEEEDVYAPLADBDKSYKQYLANGPORG 420
QY 440 RKKRYKRFMAVYDEFFKREAIQHESSGILBLYGEGDTLLIFKNQASPVYIYHGI 499
DB 421 RKKRYKRFMAVYDEFFKREAIQHESSGILBLYGEGDTLLIFKNQASPVYIYHGI 480
QY 500 TDVAPLTSRRLPKGYKHLKDPPIIPGEIFKYKTVYVEDGPTKSDPRLTRYSSFYVME 559
DB 481 TDVAPLTSRRLPKGYKHLKDPPIIPGEIFKYKTVYVEDGPTKSDPRLTRYSSFYVME 540
QY 560 RDLASGLIGPLLCIKESYDQKQIMSKDRNVIILSVYDENKSYTLLENQRLPRPAG 619
DB 541 RDLASGLIGPLLCIKESYDQKQIMSKDRNVIILSVYDENKSYTLLENQRLPRPAG 600

QY 1140 GGGSPKQVLVSIGPEKSVESQNFLEKKNVVGKGEFTKDVGLKEKVPSSBNLFTLNLD 1199
DB 763 ----- 762
QY 1200 NLHENHNOBKIOEIEKEKTELLIOENVVLPQIHVTGTGNKMKMLFLSTRONVGSY 1259
DB 763 ----- 762
QY 1260 DGATVPVLQDPRSLNSTNRKTKHTAHFSKKGEBENLEGSNQTQYVEKACTTRISPN 1319
DB 763 ----- 762
QY 1320 TSSQNFVTQSKRALKQFLPLEETELERLITVDITQNSKMKMLPSTLJQIDINEK 1379
DB 763 ----- 762
QY 1380 EKGATQSPSLSDCLTRSHSIPQANRSPPIAKVSPSPRIYLPVLEFQDNSSHLPAAS 1439
DB 763 ----- 762
QY 1440 YKKKDSGVQESSHFLQGAKKNNLSLAILLEMTGDQREVSGLSQTSATNSVYKKYENTVL 1499
DB 763 ----- 762
QY 1500 PRPDLPTKSGVELLPKVLHYOKLPEPTETNSGPHLDLVEGSLLOTEGAIKWNEANR 1559
DB 763 ----- 764
QY 1560 PGKVPFLKATESSAKTPSKLDPAMDNYGTQIPKEKWSQEKSPERTAKKKTILS 1619
DB 765 ----- 764
QY 1620 LNACENNAIAINEGSNKEIEVTWAKQGREELCSQNPVLCRQRITFTTLOSDOE 1679
DB 765 ----- 771
QY 1680 EIDYDITISVEMKEDFDIYEDENOSPSPQKTRHYFLAAVERLMDYGNSSPHVLN 1739
DB 772 KMDYDIDISTETKGEDDIYEDENOSPSPQKTRHYFLAAVEOLMDYGNSSPHVLN 831
QY 1740 RAQSGVQFKKVVQFQETGDSFTQPLVNGELNEHLGLPYTRAEVDENIWTFRNQS 1799
DB 832 RAQNGEVPFRKKVFRFADGFTQSPSYNGELNKLGLPYTRAEVDENIWTFRNQS 891
QY 1800 RPYFSYSLISYEEDQSGAEPKKNVKNPKTYEMKVOHMAPTKDECKAAYPSD 1859
DB 892 RPYFSYSLISYRDDQSGAEPKKNVKNPKTYEMKVOHMAPTKDECKAAYPSD 951
QY 1860 VDLEKDVHSLIGPLVCHTNTLNPAHQNQVTOEFALFTTIDETKSWYFTENKNCR 1919
DB 952 VDLEKDVHSLIGPLVCHTNTLNPAHQNQVTOEFALFTTIDETKSWYFTENKNCR 1011
QY 1920 APCNIQMEDPTFKENYEFHAINGYIMDLPGLVMAQDRIKWTLSMGSNENIHSIHPSG 1979
DB 1012 APCNIQMEDPTFKENYEFHAINGYIMDLPGLVMAQDRIKWTLSMGSNENIHSIHPSG 1071
QY 1980 HVEYTRKKEEKALVYLYPGVEYEMLPKSAAGIYREGLISEPHAAAMSLFLIYGSNK 2039
DB 1072 HVEYTRKKEEKALVYLYPGVEYEMLPKSAAGIYREGLISEPHAAAMSLFLIYGSNK 1131
QY 2040 CQPIGMAAGHIFDQFTASGOYQONAPKLARLHSGSINAMSTKPFEMKIDVLLPMI 2099
DB 1132 CQPIGMAAGHIFDQFTASGOYQONAPKLARLHSGSINAMSTKPFEMKIDVLLPMI 1191
QY 2100 IHGKTQGAQKQFSSSLISQFIIMYSLDGKKNQYRGNSGTGLVFFGVNDSSGIKHIF 2159
DB 1192 IHGKTQGAQKQFSSSLISQFIIMYSLDGKKNQYRGNSGTGLVFFGVNDSSGIKHIF 1251
QY 2160 NPPIIARYIRLPHYHYSIRSLTLMELMGCDLNSCMPLGMEKRAISDAQITASSFTTME 2219
DB 1252 NPPIIARYIRLPHYHYSIRSLTLMELMGCDLNSCMPLGMEKRAISDAQITASSFTTME 1311

QY 2220 ATWSPKARLHLOGRSNANRQVNNPKEMLOVDPOKMTKVTGTQGVKSLTSMYKEP 2279
DB 1312 ATWSPQARLHLOGRSNANRQVNNPKEMLOVDPOKMTKVTGTQGVKSLTSMYKEP 1371
QY 2280 LISSQDGHQWTLFFQNGKVKYQGNQDSFTPVVNSLDPLLTRLRHPOSWYQIATLR 2339
DB 1372 LVSSQDGRKMTLFLQDGHQWTLFFQNGKVKYQGNQDSFTPVVNSLDPLLTRLRHPOSWYQIATLR 1431
QY 2340 MEVLGCEADQDY 2351
DB 1432 LEVLGCEADQDY 1443
RESULT 12
US-10-115-563-14
Sequence 14, Application US/10115563
Publication No. US20030008307A1
GENERAL INFORMATION:
APPLICANT: Griffin, John H
Greengard, Judith S
TITLE OF INVENTION: METHODS FOR DIAGNOSING ACTIVATED PROTEIN
AND COMPOSITIONS THEREOF
C RESISTANCE ASSOCIATED WITH A FACTOR V GENETIC MUTATION
NUMBER OF SEQUENCES: 28
CORRESPONDENCE ADDRESS:
ADDRESS: The Scripps Research Institute, Office of
Patent Counsel
STREET: 10666 No. US20030008307A1th Torrey Pines Road, TPC 8
CITY: La Jolla
STATE: CA
COUNTRY: USA
ZIP: 92037
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/115,563
FILING DATE: 02-Apr-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/410,488
FILING DATE: 24-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Filling, Thomas
REGISTRATION NUMBER: 34,163
REFERENCE/DOCKET NUMBER: 449.0
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-554-2937
TELEFAX: 619-554-6312
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 2224 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 14:
US-10-115-563-14
Query Match 22.78; Score 2818.5; DB 9; Length 2224;
Best Local Similarity 30.48; Pred. No. 1.3e-178;
Matches 766; Conservative 394; Mismatches 847; Indels 509; Gaps 78;
QY 22 RRYLGAVALSMYDQSDGLDELVADARPPVRSFPFNTSVY-YKRTLFVEFDHLFNI 80
DB 32 RQYVAAGISMSYRBP-----PNSSLNLSTVSFKIYRYREYEPY-PRK 75
QY 81 AKRPPWAGLQPTQAEVDYDVYITIKMASHPVSLHAYGVYKMASGAYDDQDSQR 140
DB 76 EKPOSTISGLGPTLYAEVDIYKFKKADKPLSHPGQIRYKSLSGASGLDHTFPA 135
QY 141 EKEDKVPFGSHRYVWQVLKENGPMASDPLCLTYSTLSHVDLVKDLNSGLIGALLVCRE 200

136 EKMDAAVAPGREYTYEIMSISEDGPTDHPDPCLTHTIYSHENLIDEDFNSGLIGPLLCKK 195
201 GSLAKERTQ-TLHK-FILLFAVDEGKSMHSETRKNSLMODROAASARAMPKMHVNGYVA 258
196 GTLEGGTOKTFOKOYVLLFANFDESSKMSOSS-----LMTYVKNYVA 239
259 RSLPGLIGCHKRSVYVWYVGMGTTPREHSTFLEGHFVLVNRKQASLEISTITFLAQTL 318
240 GTPMDITVCADHHSIHLGMSGPELSEIHFNGOVLQONHHRVATILVSAITVSNAMT 299
319 LMDGOLFCHSHQHDGMEAVYKVDSCPEEPOLMRKNEADYDDDLTJSEMDVVA 378
300 VGPEKWIISLTPKHIQAGQAYIDINCPCRTKRLKKTRE----- 342
379 FDDNNSPFIQIRSVAKKPTWYVYIAAEBEDMDYAPVLAPDDRKYKQYLNGSPORT 438
343 -----GRHMKMEFEYIAAEVYIMDAPVIRPAMDKKTRSHLDNFNOI 387
439 GRKXKVRPMAYTDEFKTRERAI---QHEGILGPLLGEVGDTLITFKQASRPNTY 495
388 GRKXKVMYQYEDSEF-TKHTVPMNKEDGILGPIIRAOVDTLKVFKMMASRPYSY 446
496 PHGIT-----DVRPLYSRRLPKGVKILKDFPILPGEIRKMTVYEDGPTKSPCLT 549
447 PHGYTFSPYEDEVNSSF-----SGRNMTIRAVOPGETYTYKMNILEDEPTENDACLT 502
550 RYSSPVNMBERDLASGLIPLLICYKESVDORGNQIMDKRNVILFVFDENRSMYLTEN 609
503 RPYSDVIDMRDIASLIGLILICKSSLDNRGIGRAADIBOONFVNFDEKMTYLEDN 562
610 IQRLPRPAGVOLEDEFOASNMHSINGYVPSL-QLSYCHAEVAYVYIISTAQDFL 668
563 INKCEMPDEVKDDPKFESNMJSTNGYVPSITTLGFCDDIVQNHFCSTQNEIL 622
669 SVFFSGYTFKHKMYEDTLTFPESGETVPMSENPGLMILGCHNSDPRNATLKYV 728
623 TIHFTGHSFYGRHBDTLTFPMRGESEYVYTMQNVGMTLSMNSSPRSKRLKLPD 682
729 SCDKNTGDEYEDYEI-----SAYLSK--NNMIEP-----RSF 761
683 KC-----IPDDEDESEYLFEPPESTVMAFRKMHDRLEPEDESDADYDQNRLLAALGRSF 739
762 SQNRHPSSTOKOFNATITP-ENDEKTDMFAHRTPMRKIONVSSDLMLKQSTPH 820
740 RNSSL--NOEEEFNLTALENGTE-----FVSSNDIIVGSNYSSPS 781
821 GLS-----LSDLQEA--KYETFSDDPSGPAIDSNNSISEMT--HFRPOLHSGDKVFT 869
782 NISKFTVNMNLAEPOKASHQATAGSPLRHLIGKNSVLSSTAHESSP--YSEDEID 838
870 P-----ESGLQI-----RLNEKLGTAATE-----LKLIDFVYSTSNLI 905
839 PLOPDVYGIHLLSLGAFERQEHAKRGPRVEHQAAKHRSWMKLLAHVGR----- 892
906 STIPSDMLAGTNTSSLP---PSMPVHYDSOL-----DTTLFGKSSP----- 947
893 -----HLSDOTGSPGMRPWEOLPSOTGSPSPMRPWEOPSDLLLKQSNSSKILVR 946
948 ---LTESGPLSSEENNDSKLLSEGLANSOESS--WGR-----NVSSTESGFKGRAH 998
947 WHLASEKGYEITIQDDEDTAVNMNLISPOASNAWGESPLANKRGQSGHPRPVYR 1006
999 GPALLTKD--NALFKYSILKTNKTSNNSATNKXHIIDPSILLIENSPWQWILSDT 1056
1007 KSLQYRQDGGKSLKASOPLITRKKKKKKKHTH-----APLSRPFHPLRSEAY 1056
1057 EFKVATPLIDRLMDKNATLRLNHSNKTSTSKRMENYVQKKGPIPPAQN--PDMS 1114
1057 WTFSEBRRLKHSVL-----HKSNETS-----LPTLDLNTJLTSMD 1090
1115 FFKMLFLPESARWITORTHGKNSLN-SGQGPSKQOLVSLGPEKSVGQNF-LSEKNKYVG 1172

1091 FGMTIASLPD-----HNONSSNDTGOASCPGLYQYVPEE-HYQTFPIODPOK--- 1138
1173 KGEFTKDVG-----LKEMV-FPSSRRLFLVTJNDLHNHNTHNOEKI-----QEEIERK 1220
1139 --HSTDSHRSSPPLSEMLEYDSHKHSFPPTDLSQKSSSEHEWQVYISPDLSQVTL 1196
1221 ETLQENAVLPOIHVYTGKFMKMLFLLSTRONBGSYDGAIVAYLODFRSLDSTNRT 1280
1197 PELSQTNLSPLDLSHTLSPDLQRL-----SPALQOMDISPDLSH 1239
1281 ---KKTAFPSKKEEENLESGNOTQYIEKXACTYRISPNTOQNFVYORSKRALQ 1336
1240 LSPDLSH-----LSLDSQ-----TILSPDLQTNL-----SPALQ 1273
1337 FRLPEETELERLIYDOSTQMSNMKHLPS-TLQOIDYENKKGALNQSPDLQ--- 1393
1274 --MLSPDLSHHTLSDPSQTNLSPLSHMTLSPDLQTNLSP-----ALQMPISPLSH 1327
1394 -TRHSIQANRSPDLIAVSSPPIRYLYRVLQDSSHLPAASTRKKSQVQESSH 1452
1328 TTLSDPSQTNLSP-BLSQTNLSPALQMPLS-----POPSHTLSDLSQ--- 1373
1453 FLOGAKKNNLSAILTLENTQDQREVSGISATNSVYRKVENTVLP---KPLPPTSG 1509
1510 KYELLPRV-----HIYOKDLPPTENSGPGHLDY-----EGSILGTEGAIKYN 1555
1426 QMSLSPDLQVTLSPDISPTTLPLDQISPRDLOQITYPRESSQSILLO-----EPN 1479
1556 EANRGPVPLRAVATSSAKTP-----SKLDPALW-----DNHYGTQIPREKSKO 1602
1480 ES-----FPYDLOMPSPSSPLTNDQFLSEFPYLVGLSKDQTDYETIEPIEEVOSS 1534
1603 EKSPEKTAFFKKDTIISLNACSNHAIINEGONKPEIEVTYMAKQRTERLQSNPVL 1662
1535 E----- 1535
1663 KRQREITRTTQSDQEDID--YDQTSVEMKKEPDYIDDEMO--PRSPQKTRHY 1717
1536 -----DVAIEDYVPIYDPTKDYRKNINSSBDPNDIAWYLSNNGNRK 1582
1718 FLAVERLMDYKSSSPHYLNR--AOSGSVPQ---FKKYVOEFTDQSTOPLYRGELN 1772
1583 YIAEELSMY---SEFORETDIEDDDIDPEYTYKKVAYRKYLDSTPFKRPREYE 1638
1773 BHLGLPYIRAEVDINIVTFRNQASRPYSFYSLSISE-----EDROQAEPRKN 1824
1639 BHLGLPYIRAEVDYIYVRKNLASRPYSLNAGLSYEKSSGKTYEDDSPEMFREDN 1698
1825 FYKPREKTYFMYVONHMAPTDEDFCKAMAYEPDQDEKDHSLIGPLVCHTNTLP 1884
1699 AVQRRSSTYVMAHATRSQSPESGACAMAYVAYANPEKDHSLIGPLVCHTNTLP 1758
1885 AHGROVYOEPALEPTIDETKSMYFTTENMRNCRAPCNIOEDPTEKYNRPAINGY 1944
1759 DSNNEVMBREVLFTFDEKSMYWEKRSNVR-----LTSBEMKSHFPAINGMI 1812
1945 MDYLPGLVMAQDORIMYLLMSGMSNENHSIHFGHVFYRKKEEYKALNLPGVFT 2004
1813 Y-SLPGKMTQEOMVNLHLINTGSGQDIHVHFGQTLLENGRKHQDGVMPILPGSKFT 1871
2005 VEMLPKAGITRVECLIGHLIAGSSTFLVYNSKQCTPGLMAASHIDPQTISGQYQ 2064
1872 LEMKASRGWMLNTEVGENORAGQTFPLIMDRDCMPRGJSTGIDISQIKASEFLGY 1931
2065 WAPKLARLHSGSINWSTRE---PFS---WIKVDLAPMIIHGITOGAOKFSYLYIS 2118
1932 WEPLARLNNGGSVNMAVSEKILAAEFASKPMIQVDMQKVEIITGOTOGAKHYLKSCYT 1991
2119 OTIYVSLDKKNOTRGNSGTGLVWFGVNDSSGIRKHNHNEPVIATYITLHPHTHSIR 2178
1992 EYVAIYSSNQIMQIRKGNSTANVYFNGNSDASTIKENOFPPIYAKIINISPTRAYNR 2051

PRIOR FILING DATE: 1998-02-04	PRIOR APPLICATION NUMBER: 60/074086
PRIOR FILING DATE: 1998-02-09	PRIOR APPLICATION NUMBER: 60/074092
PRIOR FILING DATE: 1998-02-09	PRIOR APPLICATION NUMBER: 60/074092
PRIOR FILING DATE: 1998-02-09	PRIOR APPLICATION NUMBER: 60/077791
PRIOR FILING DATE: 1998-03-12	PRIOR APPLICATION NUMBER: 60/078910
PRIOR FILING DATE: 1998-03-20	PRIOR APPLICATION NUMBER: 60/079294
PRIOR FILING DATE: 1998-03-25	PRIOR APPLICATION NUMBER: 60/079665
PRIOR FILING DATE: 1998-02-27	PRIOR APPLICATION NUMBER: 60/079728
PRIOR FILING DATE: 1998-03-27	PRIOR APPLICATION NUMBER: 60/080165
PRIOR FILING DATE: 1998-03-31	PRIOR APPLICATION NUMBER: 60/081200
PRIOR FILING DATE: 1998-04-09	PRIOR APPLICATION NUMBER: 60/081222
PRIOR FILING DATE: 1998-04-09	PRIOR APPLICATION NUMBER: 60/081699
PRIOR FILING DATE: 1998-04-14	PRIOR APPLICATION NUMBER: 60/081817
PRIOR FILING DATE: 1998-04-15	PRIOR APPLICATION NUMBER: 60/081818
PRIOR FILING DATE: 1998-04-15	PRIOR APPLICATION NUMBER: 60/082999
PRIOR FILING DATE: 1998-04-24	PRIOR APPLICATION NUMBER: 60/083322
PRIOR FILING DATE: 1998-04-28	PRIOR APPLICATION NUMBER: 60/083545
PRIOR FILING DATE: 1998-04-29	PRIOR APPLICATION NUMBER: 60/084600
PRIOR FILING DATE: 1998-05-07	PRIOR APPLICATION NUMBER: 60/084627
PRIOR FILING DATE: 1998-05-07	PRIOR APPLICATION NUMBER: 60/084637
PRIOR FILING DATE: 1998-05-07	PRIOR APPLICATION NUMBER: 60/084637
PRIOR FILING DATE: 1998-05-07	PRIOR APPLICATION NUMBER: 60/085145
PRIOR FILING DATE: 1998-05-12	PRIOR APPLICATION NUMBER: 60/085322
PRIOR FILING DATE: 1998-05-13	PRIOR APPLICATION NUMBER: 60/085338
PRIOR FILING DATE: 1998-05-13	PRIOR APPLICATION NUMBER: 60/085338
PRIOR FILING DATE: 1998-05-13	PRIOR APPLICATION NUMBER: 60/085338
PRIOR FILING DATE: 1998-05-15	PRIOR APPLICATION NUMBER: 60/085379
PRIOR FILING DATE: 1998-05-15	PRIOR APPLICATION NUMBER: 60/085697
PRIOR FILING DATE: 1998-05-15	PRIOR APPLICATION NUMBER: 60/085704
PRIOR FILING DATE: 1998-05-15	PRIOR APPLICATION NUMBER: 60/086414
PRIOR FILING DATE: 1998-05-22	PRIOR APPLICATION NUMBER: 60/086430
PRIOR FILING DATE: 1998-05-22	PRIOR APPLICATION NUMBER: 60/086430
PRIOR FILING DATE: 1998-05-28	PRIOR APPLICATION NUMBER: 60/087106
PRIOR FILING DATE: 1998-05-28	PRIOR APPLICATION NUMBER: 60/088022
PRIOR FILING DATE: 1998-06-04	PRIOR APPLICATION NUMBER: 60/088730
PRIOR FILING DATE: 1998-06-10	PRIOR APPLICATION NUMBER: 60/088741
PRIOR FILING DATE: 1998-06-10	PRIOR APPLICATION NUMBER: 60/088810
PRIOR FILING DATE: 1998-06-10	PRIOR APPLICATION NUMBER: 60/088858
PRIOR FILING DATE: 1998-06-11	PRIOR APPLICATION NUMBER: 60/088858
PRIOR FILING DATE: 1998-06-11	PRIOR APPLICATION NUMBER: 60/089533
PRIOR FILING DATE: 1998-06-17	PRIOR APPLICATION NUMBER: 60/089599
PRIOR FILING DATE: 1998-06-17	PRIOR APPLICATION NUMBER: 60/089599


```

Db      813 LGPLNGEVDLITVYFKNMNSRFRYSVHANGVLESTVWPLAAE-----PEVVTYQM 865
Qy      1837 KYOHNAAPTKDEPDKARAYSDVLEKDVHSGLLGCPILNCHTNTLPAARGROYVQERA 1896
Db      866 NIPERSGPGNDACSMTYTSVADPKDKMTSELGAPLACQKGLIEHGRSDMDREPA 925
Qy      1897 LEFTIEDETKSMYTEMN-ERNCRAPCNOMEDPFFKENVRFHAINCYIMDTLGLVMAQ 1955
Db      926 LLEFLIEDENKSMYLEBENATGSDPSIMLQDEFPLESNKMHAINKLYANRGLTMQ 985
Qy      1956 DQIRNTILSMGSENENIHSIFSGHFTVARKKEEYKMALYNLYPGVEFTEMLPSKAGIM 2015
Db      986 GERAWMYMLAMGDDVDLHTHFHASEPLYNGENYRAADVDFPGETEVYEMVASHPGTW 1045
Qy      2016 RVECLIGENHAGKSTFLYVS-KCQTPIGMAASHI-----RDPQTASGQIGQARPL 2069
Db      1046 LMRCHYTDVHAGMETLFTVFSRTEHLSPLTVTKETEKVPPDLE-----1091
Qy      2070 ALHYSGSINAMSTKEPSPWIKRVDLLAPMTHIGIKTQAGKQKFSLYISQFIIMTSLDK 2129
Db      1092 -----EGNVYMLGMOPIPK-NVEMLASVLY-----AISVTLILVVALGCV 1131
Qy      2130 KMOTYR 2135
Db      1132 VMYQHR 1137

RESULT 15
US-10-121-049-234
: Sequence 234, Application US/10121049
: Publication No. US200302239A1
: GENERAL INFORMATION:
: APPLICANT: Baker, Kevin P.
: APPLICANT: Beresini, Maureen
: APPLICANT: DeForge, Laura
: APPLICANT: Desnoyers, Luc
: APPLICANT: Filvaroff, Ellen
: APPLICANT: Gao, Wei-Qiang
: APPLICANT: Gerritsen, Mary E.
: APPLICANT: Goddard, Audrey
: APPLICANT: Godowski, Paul J.
: APPLICANT: Gurney, Austin L.
: APPLICANT: Sherwood, Steven
: APPLICANT: Smith, Victoria
: APPLICANT: Stewart, Timothy A.
: APPLICANT: Tamas, Daniel
: APPLICANT: Watanabe, Colin K
: APPLICANT: Wood, William
: APPLICANT: Zhang, Zemin
: TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
: FILE OF INVENTION: ACIDS ENCODING THE SAME
: CURRENT APPLICATION NUMBER: US/10/121.049
: CURRENT FILING DATE: 2002-04-12
: Prior Application removed - See file wrapper or Palm
: NUMBER OF SEQ ID NOS: 550
: SEQ ID NO 234
: LENGTH: 1160
: TYPE: PRT
: ORGANISM: Homo Sapien
US-10-121-049-234

Query Match      10.7%: Score 133.5; DB 9; Length 1160;
Query local similarity 19.4%: Pred No. 5.4e-80;
Matches 421; Conservative 200; Mismatches 440; Indels 1105; Gaps 33;
Qy      20 ATRRYIAGAVELSDYI---QSDIGELPVDAREPPVPSKSF-----PENTSVYVYKTLFE 72
Db      27 ATRVYVIGIDRYOVMMNAPRGRNVITNPLDSI---VASSFLKSKNRIGGTYYKTYKE 83
Qy      73 FTDHLENIAKRPPEMGLGPTIOAEVYDVYVITIKMASHPVSLHAGVSYMRASGAE 132

```

```

Db      84 YKDSYTDVDAOPAWIGLPLVLAQEVGVILILHKNFATRPYTIHPHGVEYKDESGSL 143
Qy      133 YDQTSQREKEDDKVPGSGSHYVWVNOVLKENGPMASDPLCTYSYLSVDLKVADNSGLI 192
Db      144 YPGSSGPLKADDSVPRGSHIYNTIIPGCHAPTDADACLTLYHSHVDADRDIATGLI 203
Qy      193 GALLVCRBSL---AKETQTLHKFLLPFAVDEGKSWH-----SETKNSLMODRDA 242
Db      204 GPLITCRKALDGNPPOQROVDHDFELLESVADENLSMHLNENIATVYCSDPASVDKDE 263
Qy      243 SARAPKMHNTGVNYSRLPGLIGCHRSKYVNVHIGMTPTPEVASHLEGGHFLVNRHQ 302
Db      264 TPOESNRMAINGFVFGNLPRLANCAQKRAVNAHLMGKNEIDVTAIFHQMLTRGHHT 323
Qy      303 ASLEISPIELTAOPLMDLGOPLTLCGHSHOHGMEARVYKVSCEPEPOLMKNNBEA 362
Db      324 DVANITPATFVTAENVPEBPTMILSQVNSHFRQNALKYVCSMAPV-----375
Qy      363 EDYDDDLTDEMDVVRPDDNSPSFIOIRNSAKKHPKTVHITAAEEDMDVAPL-----417
Db      376 ----DLITG-----KROYFIEAHBIOVDYGPBGHDS 404
Qy      418 ----VLAHDRSYKSOYLNNGPORIGRKYKVFMAVYDDEFTKREAIQHS--GILGL 471
Db      405 TGRNLEBPG--SISDKFQKSSRIGGTWKVYKYEAFODETFQEKHLEEDRHILGILGV 462
Qy      472 LYGEGDPTLIIIFKNOASRPYNIYHGTIDVPRLYSRRLPGVKNLKDPLILGELTK--529
Db      463 IRAVEGDITQVYFYRNASQPSMQPHV-----FYEKDEGTV--YNDSSYVGLVAREF 515
Qy      530 ----YKWTYVDEGPTKSDPRCLTRYTSFVNMERDASGLIGLICTYESVDQRNGNI 585
Db      516 EKVYRMTYPPHAGPLADDPACLTWMTFSAADIRFVNSGLVGLVCRGALGADGKOK 575
Qy      586 MSDKRNVLLESYVDENKSMYLTENIOFLPNPAGVOL-----EDPR-FQASNMHSING 638
Db      576 GYDKEFFLETVLDEKKSYSN-----ANQAAALDPLRLESDIGFQDSNRMAING 628
Qy      639 YVPDSL-QSVCLHEVAYWYILSIGOTDPLSVFSGCYFFKKNVYEDTLTPFPSEGV 697
Db      629 FLPSNLPRLDMCKGDIYVANHLLGLCTETDVGCMRGQNTVQLCRRGAMLPHPFVNA 688
Qy      688 PMSKNPGLMILGCHNSDFRNRKGTALLKYVSCDKKGTGYDSTIEDISAYLSKNNATE 757
Db      689 IMQPDNLGTFEYICAGSHRAGRAIYVNSOC-----721
Qy      758 PRFSQNSRHPSTROKOPNATTIPENDIEKTDPMFAHRTPMKIONVSSDMLMLRQSP 817
Db      722 -----PGHQATPRORQAARI-----737
Qy      818 TPGLISLSDQEAKEYETSPDPSGALDSNNSLSEWTHFRPLQHLHSGDMVYTPSGIQLR 877
Db      738 -----737
Qy      878 LNEKLGTTAATLKLIDRFVSTSNLSLITPBDNLAACTDNTSLGPPSPVHYHQSD 937
Db      738 -----Y-----739
Qy      938 TTLPGKKSPLTESGGLSLSEBENDSKLLESGIAMSQSSGKNVSTESGILFGKRA 997
Db      740 -----IAAEVEVD-----748
Qy      998 HGPLALFKDNALFKVLSILKTNTSNNSATNKRTHIDPSLLIENSPVQNLIESDTE 1057
Db      749 -----748
Qy      1058 FKATYPLIHRMLMDKNATALRLHMSKTTSSKNMENVQKKKEPPIPDQAPNDSPFK 1117
Db      749 -----YCPDSS-----754
Qy      1118 MLPLPESARNIQRTGKNSLNSGQSPSKQVLSIGAPKESVQNFLSEKKNVYVVGGEPT 1177
Db      755 -----WEREMHNS-----EKDSYGIPLSKND-----777

```

```
QY 1178 KDVGLEKMFPPSSRLFTNLDNHEMNTNOEKKIOEIEKKETLIDENVLPQIHFTV 1237
Db 778 ----- 777
QY 1238 GTKNFMKNLFLSTQONVGSYDGAVALQDFRSLNDSTNRTKHTAFSGKEENLE 1297
Db 778 ----- 777
QY 1298 GLGNQTOIVEKACTRISPTSQNFVQBSKRALQFRLPLETELEKRIIVDDTST 1357
Db 778 ----- 777
QY 1358 QMSKNMKHLPTSTQIDYNEKEKAITQSPBDCLTRSHSIPQANRSPPLAKVSSPS 1417
Db 778 ----- 777
QY 1418 IRPIYLRVLFQDNSSHLPASGRKKDGVQSSHFLQAKKNLNLALITLMTGDQRE 1477
Db 778 ----- 777
QY 1478 VGSIGTSATNSVTYKKVENTVLPKPDLPKISGVKELLPKVHIYQKDLFPTETSNQSGHL 1537
Db 778 -GLIGS----- 782
QY 1538 DLVGSILQTEGAIKWEANRPKVPFLRVATESAKTPSKILLDPLANDNHVGTQIPKE 1597
Db 783 ----- 782
QY 1598 EKSQEKSPKTAFFKKDDTILSLNACESNHAIALINQONKPELEVYMAKQGRTERLCSQ 1657
Db 783 ----- 782
QY 1658 NPPVLKRHOREITRTLOSQDEIDYDTISVEKKEDFDIYDEENQSPRSFOKTRHY 1717
Db 783 ----- 782
QY 1718 FIAAVERLMDYGMSSPHVLNRNAGSVOPKKVFOEFTDGSFTQPLRGELNHLGL 1777
Db 783 ----- 812
QY 1778 LGPYIRAEVEDNINWTRNOASRPYSFYS-SLISYEDQOGAPRKNVYKPNETKYFW 1836
Db 813 LGPLIKGEVGDILTVFKNNASRPYSVAHGVLESTVWPLAE-----PEGVVYQW 865
QY 1837 KYQHMAPTKDEDFCKAMAYFSDVLEKDYHSLIGPLVCHTNTLPAHGRQVVOEFA 1896
Db 866 NIPERSGPPNDACSVMIIYSAVDPIKDMYSGLVPLAICOKGILEPHGRSDMDREFA 925
QY 1897 LEFTTFDETKSMYFTENK-ERNCRAPONTOMEDPTEKENYRFAHNGYIMDTLPGLVMAQ 1955
Db 926 LFLFLFEDENKSMYLEENVAHGSODPGSINLQDETFLSNKMAHNGKLYANLRLGLMTQ 985
QY 1956 DQIRWYLLSMGSNENHSHIFSGHVFYRKKEEYKMALYNLPDVEFEVEMLPKRAIWM 2015
Db 986 GERVAMYMLANGQVDLHTIHFHESFLYRNGENYRADYVDLPGEFEVEVEMVANSNPQW 1045
QY 2016 RVECLIGELHAGSSTLEFVYS-NKQOTPLGMAASHI-----ROFOITASGQYQWAPKL 2069
Db 1046 LMHCVTDDHVAHGMETLFTVFSRTEHLSPLTYTKETEKVPPROLE----- 1091
QY 2070 ARLHYSISINAWSTKEPFSWIKVDLAPMLIHGIKTOGAROKFSSLYISQFIIMYSIDK 2129
Db 1092 -----EGNVKMLGMQIPK--NVEMLASVLY-----AISVTLLLVYVLAAGV 1131
QY 2130 KQOTYR 2135
Db 1132 VWTQHR 1137
```

Search completed: July 2, 2003, 09:49:14
Job time : 110 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2003, 09:39:41 ; Search time 68 Seconds
(without alignments) 3323.706 Million cell updates/sec

Title: NP_000123

Perfect score: 12418
Sequence: 1 mglelscfcllrlfcsa.....wvhdalmevjgceagdy 2351

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

PIR_73:*
1: PIR1:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	12418	100.0	2351	1 E2HU	coagulation factor
2	8835	71.1	2319	2 A47004	coagulation factor
3	8076	65.0	2133	2 T42763	coagulation factor
4	2817.5	22.7	2224	1 KFM05	coagulation factor
5	2768.5	22.3	2183	2 T42764	coagulation factor
6	2744	22.1	2211	1 KFM05	coagulation factor
7	2190	17.6	869	2 A25945	coagulation factor
8	1320	10.6	1069	1 KTHU	ferroxidase (EC 1.11.1.1)
9	1270	10.2	1059	1 A35210	ferroxidase (EC 1.11.1.1)
10	1104	8.9	216	2 A44258	factor VIII-associated protein
11	663	5.3	427	2 JC4915	protein precursor
12	657	5.3	463	1 A36478	protein precursor
13	650	5.2	409	2 T11743	protein precursor
14	635	5.1	427	2 S63138	protein precursor
15	635	5.1	427	2 S63138	protein precursor
16	443	3.6	927	1 JQ0948	protein precursor
17	424.5	3.4	218	2 A47285	protein precursor
18	306.5	2.5	3133	2 S52093	protein precursor
19	262	2.1	845	2 JC5256	protein precursor
20	239	1.9	1072	2 A66827	protein precursor
21	236.5	1.9	3418	1 G02334	protein precursor
22	235.5	1.9	1358	2 A29363	protein precursor
23	229	1.8	2166	2 G70163	protein precursor
24	225	1.8	2401	2 T78676	protein precursor
25	221	1.8	719	2 S51739	protein precursor
26	216	1.8	2954	2 T14156	protein precursor
27	218.5	1.8	1328	2 S46773	protein precursor
28	218.5	1.8	3329	2 T30904	protein precursor
29	218.5	1.8	3507	2 T34513	protein precursor

30	217.5	1.8	3329	2 T42205	breast cancer susc
31	217	1.7	1381	1 S45781	probable calcium-b
32	216.5	1.7	1271	2 D64237	hypothetical prote
33	213.5	1.7	2346	2 T13829	hypothetical prote
34	213.5	1.7	3328	2 T30835	hypothetical prote
35	213.5	1.7	5005	2 F82884	hypothetical prote
36	212	1.7	1957	2 T38077	hypothetical colle
37	212	1.7	3924	2 S37431	ankyrin 2, neurom
38	211	1.7	1283	2 T13799	neurexin IV - fru
39	209.5	1.7	1177	2 T64233	hypothetical prote
40	209	1.7	1420	1 A44361	hypothetical prote
41	208.5	1.7	1628	2 E90538	hypothetical prote
42	207.5	1.7	737	2 T31249	hypothetical prote
43	206.5	1.7	737	2 T13615	hypothetical prote
44	204.5	1.6	1805	2 T02712	similar to late em
45	204	1.6	1165	2 S62982	vacuolar protein V

ALIGNMENTS

RESULT 1

coagulation factor VIII precursor [validated] - human
E2HU
N:Alternate names: antihemophilic factor A; coagulation factor VIII; procoagular
C:Species: Homo sapiens (man)
C:Date: 26-Aug-1985 #sequence-revision 28-Aug-1985 #text-change 08-Dec-2000
C:Accession: I54318; A00525; I58059; A23584; A26174; A42348; A43986; S63527; S664
R:Gitschler, J.; Wood, W.I.
Hum. Mol. Genet. 1, 199-200, 1992
A:Title: Sequence of the exon-containing regions of the human factor VIII gene.
A:Reference number: I54318; MUID:93255012; PMID:1303178
A:Accession: I54318
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1921, 'S', 1923-2351 <RES>
A:Cross-references: GB:M68648; NID:918281; PIDN:AA52420.1; PID:918281
R:Wood, W.I.; Capon, D.J.; Simonsen, C.C.; Eaton, D.L.; Gitschler, J.; Keyt, B.;
Nature 312, 330-337, 1984
A:Title: Expression of active human factor VIII from recombinant DNA clones.
A:Reference number: A00525; MUID:85061548; PMID:6438526
A:Accession: A00525
A:Molecule type: mRNA
A:Residues: 1-2351 <MO>
A:Cross-references: EMBL:X01165; EMBL:X01166; EMBL:X01179
R:Toole, J.J.; Knopf, J.L.; Wozney, J.M.; Sultzman, L.A.; Buecker, J.L.; Pittman,
S.; D.N.; Hewick, R.M.
Nature 312, 342-347, 1984
A:Title: Molecular cloning of a cDNA encoding human antihemophilic factor.
A:Reference number: I58059; MUID:85061550; PMID:6438528
A:Accession: I58059
A:Status: nucleic acid sequence not shown; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-74, 'V', 76-1259, 'E', 1261-2351 <RES>
A:Cross-references: GB:X01740; NID:9182802; PIDN:AA52484.1; PID:9182803
R:Rittett, M.A.; Blacher, R.; Burke, R.L.; Caput, D.; Chu, C.; Dima, D.; Hartog, K.
B.; Randolph, A.; Orde, M.S.; Valenzuela, P.; Dahl, H.H.; Favalaro, J.; Hansen,
DNA 4, 333-349, 1985
A:Title: Characterization of the polypeptide composition of human factor VIII:C at
A:Reference number: A23584; MUID:86081164; PMID:3935400
A:Accession: A23584
A:Molecule type: mRNA
A:Residues: 1-2351 <TR>
A:Cross-references: GB:M4113; NID:9182817; PIDN:AA52485.1; PID:9182818
R:Eaton, D.; Rodriguez, H.; Vohar, G.A.
Biochemistry 25, 505-512, 1986
A:Title: Proteolytic processing of human factor VIII. Correlation of specific clea
A:Reference number: A26174; MUID:86159740; PMID:3082357
A:Accession: A26174
A:Molecule type: protein
A:Residues: 20-36, 392-399, 'X', 401-402, 1668-1678, 1709-1722, 'D', 1723-1725, 1741-1755
R:Pittman, D.D.; Wang, J.H.; Kaufman, R.J.

	Query Match	100.0%;	Score 12418;	DB 1;	Length 2351;
	Best Local Similarity	100.0%;	Pred. No. 0;		
	Matches 2351;	Conservative 0;	Mismatches	Indels 0;	Gaps 0;
QY	1 MOELSTCFEFLCLRFCSFATRRYYLGAVALSLMDYMSDGLGELPVDARPPRPVKSPFFN	60			
DB	1 MOELSTCFEFLCLRFCSFATRRYYLGAVALSLMDYMSDGLGELPVDARPPRPVKSPFFN	60			
QY	61 TSYVYKRLTFEFDHFLFNAKRPPRMGLGPTLOAEYDVAVITLTKMAHPSLAV	120			
DB	61 TSYVYKRLTFEFDHFLFNAKRPPRMGLGPTLOAEYDVAVITLTKMAHPSLAV	120			
QY	121 GYSYKRAEAGAEYDDQSOBEKEDKVPFGSGHYYWQYLKENGPAASPLCLTYSTLH	180			
DB	121 GYSYKRAEAGAEYDDQSOBEKEDKVPFGSGHYYWQYLKENGPAASPLCLTYSTLH	180			
QY	181 VDLVYDNLNSGLIGALLVCRGSLAKKKTOTLHKLFLPAVFDGKSWSEYTNISLMODRD	240			
DB	181 VDLVYDNLNSGLIGALLVCRGSLAKKKTOTLHKLFLPAVFDGKSWSEYTNISLMODRD	240			
QY	241 AASARAPRMATYGVNGYVNSRLPGLICCHRSYVMYVIGMGTPEVHSIFLEGHTELVNHN	300			
DB	241 AASARAPRMATYGVNGYVNSRLPGLICCHRSYVMYVIGMGTPEVHSIFLEGHTELVNHN	300			
QY	301 ROASLEISPIEFLAQTLLMDLGOFLFLCHISRHQDMEXYVAVDSCEEPQLMKRKNNE	360			
DB	301 ROASLEISPIEFLAQTLLMDLGOFLFLCHISRHQDMEXYVAVDSCEEPQLMKRKNNE	360			
QY	361 EAEYDYYDOLTSBMDYVRFDDNSPFIOTRSVAKKHPTKMYVYIAAEEDMDYAPLVLA	420			
DB	361 EAEYDYYDOLTSBMDYVRFDDNSPFIOTRSVAKKHPTKMYVYIAAEEDMDYAPLVLA	420			
QY	421 PDDRSYKSGYTLNNGPQRIKGRKKYKFAFYDTETKTRREALQHSIGLPLLYGEGVTL	480			
DB	421 PDDRSYKSGYTLNNGPQRIKGRKKYKFAFYDTETKTRREALQHSIGLPLLYGEGVTL	480			
QY	481 LIIFKQASRPYNTYPHGITDVAPRLYSRRLPGVXHLKDPFLIYGEIIFYKMTVYEDGP	540			
DB	481 LIIFKQASRPYNTYPHGITDVAPRLYSRRLPGVXHLKDPFLIYGEIIFYKMTVYEDGP	540			
QY	541 TKSDPCLCLTRYSSFYVNNERDLASGLIGLPLLCYSESYDQGNQINMSDKRNLYLSVDE	600			
DB	541 TKSDPCLCLTRYSSFYVNNERDLASGLIGLPLLCYSESYDQGNQINMSDKRNLYLSVDE	600			
QY	601 NRSWYLTENIOFLPNPAGVOLDPEFQASNMHSINCYVDSJOLSYCLHEVAWYTLIS	660			
DB	601 NRSWYLTENIOFLPNPAGVOLDPEFQASNMHSINCYVDSJOLSYCLHEVAWYTLIS	660			
QY	661 IGAQDQDFLSVFFSGYTRKHKMYEDTLTLFPSSGGTYVMSNENGLMILGCHNSDFRRNG	720			
DB	661 IGAQDQDFLSVFFSGYTRKHKMYEDTLTLFPSSGGTYVMSNENGLMILGCHNSDFRRNG	720			
QY	721 MPALILVSSCDKNGDYEDSYEDISAYLLSKNNAIEBSPSONSRHPSTROKQFNATTI	780			
DB	721 MPALILVSSCDKNGDYEDSYEDISAYLLSKNNAIEBSPSONSRHPSTROKQFNATTI	780			
QY	781 PEADIKETDPMFAHHTPMKIQONVSSDMLMLRSPPHLSJLSDJQEKYETFECDPS	840			
DB	781 PEADIKETDPMFAHHTPMKIQONVSSDMLMLRSPPHLSJLSDJQEKYETFECDPS	840			
QY	841 PGALDSNNSLSMTHFRQLHSGDWYTFPESGLOLRNKLGLTTATTELAKLIDFVYSST	900			
DB	841 PGALDSNNSLSMTHFRQLHSGDWYTFPESGLOLRNKLGLTTATTELAKLIDFVYSST	900			

Db	841	PGALDSNNLSSEHTFHRRQJLHSGDMYF7PESQLOJLRLNEKLGTTAATLKLKIDRKYST	900
Oy	901	SNNLISITPBDNIAACTDNTSSICLPBSMAYHDSLOLTTFLFGKSSPLTESGCVLSJEE	960
Db	901	SNNLISITPBDNIAACTDNTSSICLPBSMAYHDSLOLTTFLFGKSSPLTESGCVLSJEE	960
Oy	961	NNDSKLLSEGLMANSQSSSGKKNVSTSGRLTFGKGRAGPALLRKDUALFKVYSISLSTN	1020
Db	961	NNDSKLLSEGLMANSQSSSGKKNVSTSGRLTFGKGRAGPALLRKDUALFKVYSISLSTN	1020
Oy	1021	KTSNNKSNKKTHTDPSLITENSPSWOMLIESOPFERKTYPLIHRMAMOKNMTAPL	1080
Db	1021	KTSNNKSNKKTHTDPSLITENSPSWOMLIESDTERKTYPLIHRMAMOKNMTAPL	1080
Oy	1081	NHMSKNTTSSKNEMVQOKKEGPIPPDAQNPDMKSFKMLFLPESARMIQTHCKNSLNSG	1140
Db	1081	NHMSKNTTSSKNEMVQOKKEGPIPPDAQNPDMKSFKMLFLPESARMIQTHCKNSLNSG	1140
Oy	1141	QGPSPKQVLSIGPERVEKQNFLEKKNVYVYGGETTKVQGLKEMVPPSRNLFITLND	1200
Db	1141	QGPSPKQVLSIGPERVEKQNFLEKKNVYVYGGETTKVQGLKEMVPPSRNLFITLND	1200
Oy	1201	LHENNTAHNOEKTOEIEEKKETPLQIOWENVLPQIHYTGKFKPKNFLELSTQNWBGSTD	1260
Db	1201	LHENNTAHNOEKTOEIEEKKETPLQIOWENVLPQIHYTGKFKPKNFLELSTQNWBGSTD	1260
Oy	1261	GAYAPVLODPRSLNDSTNTKKTAAHFSKSGKEENLEGJNOQTOIYEKTYACTTRISPPT	1320
Db	1261	GAYAPVLODPRSLNDSTNTKKTAAHFSKSGKEENLEGJNOQTOIYEKTYACTTRISPPT	1320
Oy	1321	SOQNPVYQSRKALKQFRLPLEETELKRIIYVDTSTOKSMKMHLPSTLQIDVNEKE	1380
Db	1321	SOQNPVYQSRKALKQFRLPLEETELKRIIYVDTSTOKSMKMHLPSTLQIDVNEKE	1380
Oy	1381	KCATQSPFLSDCLTRSHSTIPQANSPLIAKVSFSPSTIRPLYTRVLFDONSHLPASAY	1440
Db	1381	KCATQSPFLSDCLTRSHSTIPQANSPLIAKVSFSPSTIRPLYTRVLFDONSHLPASAY	1440
Oy	1441	RKQSGVOESHPLQOAKKNLISLAITLPMTDQDREVSIGTATSATSVYKKKEVNTVLP	1500
Db	1441	RKQSGVOESHPLQOAKKNLISLAITLPMTDQDREVSIGTATSATSVYKKKEVNTVLP	1500
Oy	1501	KRDLPKTSQVAYELLPVNHYXQDLEPPEETNSGPHGLDYBSLLOGTBEAKIWMANP	1560
Db	1501	KRDLPKTSQVAYELLPVNHYXQDLEPPEETNSGPHGLDYBSLLOGTBEAKIWMANP	1560
Oy	1561	GVPPFLRYATPSSAKTPPSKLDDPLAMDNHYGTOIPREEMKSOEKSPKTAFFKKKDYIISL	1620
Db	1561	GVPPFLRYATPSSAKTPPSKLDDPLAMDNHYGTOIPREEMKSOEKSPKTAFFKKKDYIISL	1620
Oy	1621	NACSESNHAIAMIEGONKREIEVTMAQOGFERLKQSNPVLKRRHOEIRRTTLQSDOE	1680
Db	1621	NACSESNHAIAMIEGONKREIEVTMAQOGFERLKQSNPVLKRRHOEIRRTTLQSDOE	1680
Oy	1681	IDYDUTTSYEMKKEEDDIYDEBENSPPRSOKTBYFTAAVERLIMYGSSSPHYLNR	1740
Db	1681	IDYDUTTSYEMKKEEDDIYDEBENSPPRSOKTBYFTAAVERLIMYGSSSPHYLNR	1740
Oy	1741	AQSGSVPOFKKVVFOEFTDGSFTOPLYKGBELNHLGLLGPYIRAEVDNIMTFRNQAS	1800
Db	1741	AQSGSVPOFKKVVFOEFTDGSFTOPLYKGBELNHLGLLGPYIRAEVDNIMTFRNQAS	1800
Oy	1801	PSFSPSSLISYEEDQOQAEPRKNVYKNETKITVFKVQVHNHAPKQEPQCKAMAAFSYV	1860
Db	1801	PSFSPSSLISYEEDQOQAEPRKNVYKNETKITVFKVQVHNHAPKQEPQCKAMAAFSYV	1860
Oy	1861	DLERKXVHGLIGLPLACHNTLNPANHGROYTOERALFTTIDEXKSMYTFEMNENCA	1920
Db	1861	DLERKXVHGLIGLPLACHNTLNPANHGROYTOERALFTTIDEXKSMYTFEMNENCA	1920
Oy	1921	PCNIOMEDPFFKEMVRFHAINGYIMDULPGIYAAQODRIRWLLLSMGSNENIISIFHSQ	1980
Db	1921	PCNIOMEDPFFKEMVRFHAINGYIMDULPGIYAAQODRIRWLLLSMGSNENIISIFHSQ	1980

QY	1981	VETVARKREYEMALYNLYGVEYETEMLPKSAIGIMRECELTIGEHILHAGKSTLFLYVSNKC	2040
DB	1981	VETVARKREYEMALYNLYGVEYETEMLPKSAIGIMRECELTIGEHILHAGKSTLFLYVSNKC	2040
QY	2041	QFFPLMASGHIRODTQTASQIGQNAAPKLARLHSSGINASTKEPFSIKYDLAPMT	2100
DB	2041	QFFPLMASGHIRODTQTASQIGQNAAPKLARLHSSGINASTKEPFSIKYDLAPMT	2100
QY	2101	HGIKTQGAOKFSSLYISOFIIMYSLSDCKKQOTYRGNSTGTLWVEFGVNDSSGIKRHINFN	2160
DB	2101	HGIKTQGAOKFSSLYISOFIIMYSLSDCKKQOTYRGNSTGTLWVEFGVNDSSGIKRHINFN	2160
QY	2161	PELIRARYRLPHPTVYSIRSTRLMEWELGCDLNSCAMPKGMESKASIDAOITASSEFTNMA	2220
DB	2161	PELIRARYRLPHPTVYSIRSTRLMEWELGCDLNSCAMPKGMESKASIDAOITASSEFTNMA	2220
QY	2221	TWSPKARLHIGGSNMRQVNNPKNPEMLQVDPKTKMVGNTGQVRSILTSMYVEFL	2280
DB	2221	TWSPKARLHIGGSNMRQVNNPKNPEMLQVDPKTKMVGNTGQVRSILTSMYVEFL	2280
QY	2281	ISSSDGHOMTLFFONGKAVAFEGNDSFTPVNSLDPPLTFLRLIHPQSNVHOLALRM	2340
DB	2281	ISSSDGHOMTLFFONGKAVAFEGNDSFTPVNSLDPPLTFLRLIHPQSNVHOLALRM	2340
QY	2341	EVLGCEADLY 2351	
DB	2341	EVLGCEADLY 2351	

RESULT 2

A47004
 coagulation factor VIII precursor - mouse
 C:Species: Mus musculus (house mouse)
 C:Date: 09-Sep-1994 #sequence, revision 09-Sep-1994 #text, change 18-Jun-1999
 G:Accession: A47004
 R:Eider, B.; Larkich, D.; Glitschler, J.
 Genomics 16, 374-379, 1993
 A>Title: Sequence of the murine factor VIII cDNA.
 A:Reference number: A47004; M01D:93300511; PMID:8314577
 A:Accession: A47004
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-2319 <EID>
 A:Cross-references: GB:105573; NID:g192456; PIDN:AA37385.1; PID:g192457
 G:Superfamily: coagulation factor VIII; discoidin 1 amino-terminal homology; ferritin
 F:1-19/Domain: signal sequence #status predicted <SIG>
 F:23-349/Domain: ferroxidase repeat homology <FO>
 F:402-730/Domain: ferroxidase repeat homology <RO>
 F:1866-2006/Domain: ferroxidase repeat homology <RO>
 F:2007-2136/Domain: discoidin 1 amino-terminal homology <DN1>
 F:2160-2313/Domain: discoidin 1 amino-terminal homology <DN>

Query Match 71.1% Score 8835; DB 2; Length 2319;
 Best Local Similarity 72.5% Pred. No. 0;
 Matches 1713; Conservative 232; Mismatches 363; Indels 54; Gaps 22;

QY	1	MOLEISTCFEFLCLRFECFSATRRYVYLGAVELSWDYMOSD-LGELPYDARFPKPRVKSPEF	59
DB	1	MOLEISTCFEFLCLRFECFSATRRYVYLGAVELSWDYMOSD-LGELPYDARFPKPRVKSPEF	60
QY	60	NNSVYKKTLPEVETDLFINAKPRPPMNGILGPIIAEYDVRVYITLNNASHPVSIHA	119
DB	61	NNSVYKKTLPEVETDLFINAKPRPPMNGILGPIIAEYDVRVYITLNNASHPVSIHA	120
QY	120	VGVSTWKAASEAIEDDOTSQREKEDKRTFGSGSRYYVQVLKENGPMASDPCLTSTLS	179
DB	121	VGVSTWKAASEAIEDDOTSQREKEDKRTFGSGSRYYVQVLKENGPMASDPCLTSTLS	180
QY	180	HDVLVNDLNSGLIGALLVCREGSLAREKTQTLHKEFILLFAVDEGRSWHSETRKNSLMDR	239
DB	181	HDVLVNDLNSGLIGALLVCREGSLAREKTQTLHKEFILLFAVDEGRSWHSETRKNSLMDR	240

QY	240	DASABRAMEKMTYNGVYNSJLPGLLGCRKSYVNYHNGTTPRHNSTFLEGTETLVN	299
Db	241	DSASABRMEKMTYNGVYNSJLPGLLGCRKSYVNYHNGTTPRHNSTFLEGTETLVN	300
QY	300	HROSLEIPEFLEAQTLLMDLGOFLERCHLSSHQHDGEAVYKVDSPCEPOLRKN-N	358
Db	301	HROSLEIPEFLEAQTLLMDLGOFLERCHLSSHQHDGEAVYKVDSPCEPOLRKN-N	360
QY	359	NEKEADYDDDLDSMDVPRDDONSPTQJRSYAKNHPYVWHYIAEEDMDYAPLV	418
Db	361	NEKEADYDDDL-SMDVFTLDDSP-FIQJRSYAKNHPYVWHYIAEEDMDYAPSV	418
QY	419	LAPDNRYSQYLNNGPORIGRKYKRYEMATDTFTFRALIQHSSGLTGLXGEVD	478
Db	419	PTSDNGSYKQYLSNGPBRIGRKYKRYEIAATDFTFTFRALQHSGLTGLXGEVD	478
QY	479	TLLIIFNQAQSRPYNYPHGITDVRPLYXSRNLKPKGVKTLKDFLPGELFKYKMTYVED	538
Db	479	TLLIIFNQAQSRPYNYPHGITDVRPLYXSRNLKPKGVKTLKDFLPGELFKYKMTYVED	538
QY	539	GPTSDRCITRYYSFPMNMRDLASGLIGPLITCYKSYOQRONQJMDKRNVLTFVSF	598
Db	539	GPTSDRCITRYYSFPIPERDLASGLIOPILITCKSYOQRONQJMDKRNVLTFVSIF	598
QY	599	DENRSWYLTENIORFLPNPAGVLEDEPEFQASNMHSTNGYVDSIQLSYCHAEVAYYI	658
Db	599	DENRSWYLTENIORFLPNPAAQTQPODQFQASNMHSTNGYVDSIELYVCHAEVAYYI	658
QY	659	LSIGAQNDPFLVSFGSYCFKRMVYEDTLLTFPFGSETVFMSMBPGLWILCHNSDERN	718
Db	659	LSVGAQNDPFLVSFGSYCFKRMVYEDTLLTFPFGSETVFMSMBPGLWILCHNSDERK	718
QY	719	RGMRLALVYSCCKNGDYEDSEYEDISAVLLSKNNALTEPFSQNSRHPSTROKOPAT	778
Db	719	RGMRLALVYSCCKNSDYEDIEYEDIPQLVNNENVYIDPSPFQNNHNPTRKKFKFDS	778
QY	779	TTPENDLEKTDPMFARITPMKIQNVSSDMLLRQS-PPHPLSTSDIOEKYETSD	837
Db	779	TTPKMDMEKTEPEFEBIAEMKLVSVSQMLMLQOSHPHGLFSLDQOEALIBAHID	838
QY	838	DSPGDAIDSNNSLSEMTWHRPQLNHSQDMVFTPESGLQRLNKEIKGTYAATELKIDRFY	897
Db	839	DSPDAIDSNNGSPSKYQJLRPSHSHSEKIEYFTQPEQJLQJNSKSLIETIIVYKKKJGLQY	898
QY	898	SSTSNMLI-STIPSDNLAGTONTNSSLGPPSMVYVHDSDQJLTFFGKSSSPLTESGQPLS	956
Db	899	SSIPENLMTTIIISDLNKAFTFKDSSQFDMVYHSSKLSITAFGKKAISLVGSHVPLN	958
QY	957	LSEENDSKLIESGLNMOESSGWMGKANSSTBSGRLEFKGRAHPALLTKONALFKYSISL	1016
Db	959	ASEENSDSNIIIDSTLWYSOESLPRNIIJLSEIBNRLLRERFRFGALLTKONTLEKONSL	1018
QY	1017	LKTNKTSNNSATNKKHIGDPLLENSPSWQV-ILSEBDTEFFKAVPLIHDMLMDKNA	1075
Db	1019	MKTNTYVHSTNKKLHTSEPT-SJENSTTDLOALIKNSBIECVYALJHDSTTLGKNS	1077
QY	1076	TALRLNHSKNTYSKKMMEVOOKKEGPIPPDQNDQMSFPMKLPERSARWIOHTGKN	1135
Db	1078	TYLRNHLNMTYSTKAKDIFHRKQEDPIPODEBNTIIPFSMLPLSSSSWFKKTKNGN	1137
QY	1136	SLNSGQGPSPOLVSLGPEYVEGONLSEKKNVYVGKEPTKDYGLKEMVFPSSNLEL	1195
Db	1138	SLNSEQESHPOLVYLTFKKYVKNQSPLESEKNKYVEOGFTKNGLKDMAFPHNSISPL	1197
QY	1196	TYNQDNLHENHNOEKIOEIEKEKTELIOENVVLPOITVGTGTFENFKMLISTRQV	1255
Db	1198	TITLSNHHNGHNEKNOIEEIE-KEALIEKVLPQVDEHATGSKNFLDILIGTRQNI	1256
QY	1256	EGSYGDAIAPVLODFRSLDSTNTRKKTAFKRS--KGEENMLDEIGMOPQVLEYKACT	1313
Db	1257	-SLEVHPVPAQNTISJNSTVYQIIMHEFFRRKDKINSBQJLKEVANKY----	1311
QY	1314	TRISPTSQONFVQJRSKRLKQFRLPELETLEKRIIYDNTSTQWSKNMKNHLPSTLQ	1373

Db	1312	-----P\$Q\$NTITQ\$KRALG\$FRL-----ST\$M\$K\$T\$INC\$T\$OC\$IIKO	1349
Qy	1374	IDV\$E\$K\$K\$A\$T\$O\$P\$L\$O\$C\$T\$R\$H\$S\$I\$Q\$A\$N\$R\$P\$L\$A\$Y\$S\$P\$P\$R\$T\$Y\$T\$R\$V\$L\$F\$O\$D\$N\$S	1433
Db	1350	IDH\$E\$K\$E\$K\$A\$T\$T\$S\$S\$D\$S--\$V\$Y\$S\$T\$O\$T\$M\$S\$D\$H\$Y\$K\$T\$S\$A\$P\$---P\$D\$K\$R\$P\$O\$N\$K\$S	1405
Qy	1434	HL\$P\$A\$Y\$---\$R\$K\$D\$S\$V\$O\$E\$S\$H\$P\$L\$O\$A\$K\$K\$N\$N\$S\$A\$T\$L\$E\$K\$T\$G\$D\$Q\$E\$V\$S\$D\$T\$S\$A\$T\$N\$Y	1468
Db	1406	H\$V\$O\$A\$S\$S\$Y\$T\$D\$F\$K\$T\$R\$S\$R\$I\$O\$E\$S\$N\$F\$L\$E\$K\$T\$K\$N\$P\$S\$A\$T\$L\$E\$P\$M\$N\$F\$ID\$O\$K\$F\$S\$P\$K\$S\$T\$N\$Y	1465
Qy	1490	T\$Y\$K\$E\$V\$T\$V\$P\$R\$D\$E\$T\$K\$T\$G\$K\$V\$E\$LL\$P\$V\$H\$Y\$Q\$D\$E\$T\$P\$T\$S\$N\$G\$P\$G\$H\$D\$U\$E\$G\$S\$L\$Q\$G\$T\$E	1549
Db	1466	T\$Y\$K\$R\$E\$N\$I\$F\$L\$K\$T\$P\$E\$E\$G\$K\$T\$E\$LL\$P\$O\$V\$S\$I\$O\$E\$B\$E\$I\$P\$E\$T\$S\$H\$G\$P\$G\$H\$N\$A\$K\$E\$V\$F\$Q\$K\$IO	1525
Qy	1550	G\$A\$K\$M\$N\$A\$N\$R\$G\$V\$P\$P\$L\$A\$Y\$A\$T\$E\$S\$A\$K\$P\$K\$P\$S\$K\$L\$D\$P\$L\$A\$M\$D\$H\$Y\$O\$T\$P\$K\$E\$M\$K\$O\$E\$K\$P\$E\$K\$T	1609
Db	1526	G\$T\$T\$M\$N\$A\$K\$K\$R\$G\$E--\$I\$K\$T\$E\$S\$S\$A\$N\$K\$R\$K\$L\$N\$H\$M\$D\$H\$Y\$A\$Q\$P\$D\$M\$K\$S\$K\$E\$K\$P\$E\$I	1583
Qy	1610	A\$E\$K\$K\$D\$T\$T\$S\$L\$N\$C\$E\$S\$N\$H\$A\$A\$A\$N\$E\$G\$K\$P\$E\$V\$E\$Y\$A\$K\$O\$G\$T\$E\$R\$L\$C\$O\$N\$P\$V\$Y\$K\$K\$Q\$E\$B\$I	1665
Db	1584	S\$I\$K\$O\$E\$T\$T\$L\$R\$P\$H\$G\$N\$S\$H\$G\$A--\$N\$E\$K\$O\$M\$P\$R\$E\$T\$Y\$W\$O\$G\$O\$F\$O\$T\$C\$Q\$D\$P\$V\$Y\$K\$R\$Q\$R\$E\$L	1647
Qy	1670	T\$E\$T\$T\$O\$S\$O\$E\$B\$Y\$D\$D\$T\$I\$E\$V\$E\$K\$K\$D\$E\$P\$Y\$D\$E\$D\$E\$N\$O\$S\$P\$R\$F\$O\$K\$T\$R\$H\$E\$T\$A\$V\$E\$R\$L\$M\$D\$G	1722
Db	1643	--\$A\$N\$O\$Q\$E\$A\$T\$Y\$D\$A\$T\$E--T\$I\$D\$P\$Y\$S\$E\$D\$T\$K\$O\$G\$P\$S\$F\$O\$K\$T\$R\$H\$E\$T\$A\$V\$E\$R\$L\$M\$D\$G	1695
Qy	1730	M\$E\$S\$S\$H\$Y\$T\$A\$N\$A\$O\$G\$S\$V\$P\$O\$K\$K\$V\$Y\$O\$E\$T\$O\$S\$P\$O\$P\$Y\$G\$E\$L\$N\$E\$H\$G\$L\$G\$Y\$T\$A\$E\$V\$E\$N	1789
Db	1700	M\$E\$T\$S--\$H\$V\$R\$N\$T\$O\$G\$D\$N\$P\$O\$K\$K\$V\$Y\$O\$E\$T\$D\$S\$F\$O\$P\$Y\$G\$E\$L\$N\$E\$H\$G\$L\$G\$Y\$T\$A\$E\$V\$E\$N	1755
Qy	1790	I\$M\$V\$E\$R\$N\$O\$A\$S\$P\$R\$Y\$S\$S\$T\$S\$E\$D\$O\$G\$A\$P\$R\$K\$N\$Y\$K\$P\$R\$E\$K\$T\$Y\$F\$K\$V\$O\$H\$N\$A\$P\$T\$D\$E\$F	1844
Db	1759	I\$M\$V\$E\$T\$K\$N\$O\$A\$S\$P\$R\$Y\$S\$S\$T\$S\$E\$D\$O\$G\$A\$P\$R\$K\$N\$Y\$K\$P\$R\$E\$K\$T\$Y\$F\$K\$V\$O\$H\$N\$A\$P\$T\$D\$E\$F	1811
Qy	1850	D\$K\$A\$N\$A\$Y\$S\$D\$V\$D\$E\$K\$Y\$H\$S\$G\$L\$G\$P\$L\$V\$O\$H\$T\$L\$T\$P\$A\$H\$G\$R\$O\$Y\$T\$V\$O\$E\$F\$A\$F\$T\$I\$D\$E\$T\$K\$S\$W\$T	1909
Db	1818	D\$K\$A\$N\$A\$Y\$S\$D\$V\$D\$E\$K\$Y\$H\$S\$G\$L\$G\$P\$L\$V\$O\$H\$T\$L\$T\$P\$A\$H\$G\$R\$O\$Y\$T\$V\$O\$E\$F\$A\$F\$T\$I\$D\$E\$T\$K\$S\$W\$T	1877
Qy	1910	F\$E\$N\$M\$E\$R\$C\$A\$P\$C\$I\$O\$M\$E\$D\$P\$T\$E\$K\$N\$R\$H\$A\$N\$G\$Y\$M\$D\$T\$P\$G\$Y\$M\$A\$O\$D\$R\$I\$M\$Y\$T\$L\$S\$G\$N\$S	1961
Db	1878	F\$E\$N\$M\$E\$R\$C\$A\$P\$C\$I\$O\$M\$E\$D\$P\$T\$E\$K\$N\$R\$H\$A\$N\$G\$Y\$M\$D\$T\$P\$G\$Y\$M\$A\$O\$D\$R\$I\$M\$Y\$T\$L\$S\$G\$N\$S	1933
Qy	1970	E\$N\$H\$S\$H\$E\$S\$G\$H\$E\$T\$V\$R\$K\$E\$E\$Y\$K\$M\$A\$Y\$N\$T\$P\$G\$V\$E\$T\$Y\$E\$M\$P\$S\$K\$A\$G\$T\$R\$V\$E\$B\$C\$L\$G\$H\$A\$N\$G	2022
Db	1938	E\$N\$H\$S\$H\$E\$S\$G\$H\$E\$T\$V\$R\$K\$E\$E\$Y\$K\$M\$A\$Y\$N\$T\$P\$G\$V\$E\$T\$Y\$E\$M\$P\$S\$K\$A\$G\$T\$R\$V\$E\$B\$C\$L\$G\$H\$A\$N\$G	1999
Qy	2030	S\$T\$E\$F\$Y\$S\$N\$K\$O\$P\$T\$P\$G\$A\$S\$H\$I\$N\$D\$F\$O\$T\$A\$S\$O\$Y\$O\$M\$P\$K\$A\$R\$A\$H\$Y\$S\$G\$S\$I\$N\$A\$M\$S\$T\$R\$E\$P\$S\$W	2088
Db	1998	S\$T\$E\$F\$Y\$S\$N\$K\$O\$P\$T\$P\$G\$A\$S\$H\$I\$N\$D\$F\$O\$T\$A\$S\$O\$Y\$O\$M\$P\$K\$A\$R\$A\$H\$Y\$S\$G\$S\$I\$N\$A\$M\$S\$T\$R\$E\$P\$S\$W	2055
Qy	2090	I\$K\$Y\$D\$L\$A\$P\$H\$T\$H\$G\$T\$O\$G\$A\$Q\$K\$T\$S\$S\$Y\$S\$O\$P\$T\$M\$S\$D\$G\$K\$M\$O\$T\$Y\$G\$N\$S\$T\$G\$T\$A\$W\$E\$P\$G\$N	214
Db	2058	I\$K\$Y\$D\$L\$A\$P\$H\$T\$H\$G\$T\$O\$G\$A\$Q\$K\$T\$S\$S\$Y\$S\$O\$P\$T\$M\$S\$D\$G\$K\$M\$O\$T\$Y\$G\$N\$S\$T\$G\$T\$A\$W\$E\$P\$G\$N	211
Qy	2150	D\$S\$G\$I\$K\$H\$N\$F\$N\$P\$P\$T\$A\$R\$Y\$R\$L\$H\$P\$H\$S\$I\$N\$S\$T\$L\$R\$E\$L\$G\$C\$D\$N\$S\$C\$M\$P\$L\$E\$S\$K\$A\$S\$D\$A\$O\$I	220
Db	2118	D\$S\$G\$I\$K\$H\$N\$F\$N\$P\$P\$T\$A\$R\$Y\$R\$L\$H\$P\$H\$S\$I\$N\$S\$T\$L\$R\$E\$L\$G\$C\$D\$N\$S\$C\$M\$P\$L\$E\$S\$K\$A\$S\$D\$A\$O\$I	217
Qy	2210	T\$A\$S\$Y\$T\$N\$E\$A\$T\$W\$P\$S\$A\$R\$L\$H\$O\$G\$R\$N\$A\$R\$P\$V\$N\$P\$K\$E\$M\$O\$V\$O\$F\$O\$T\$M\$K\$V\$T\$S\$V\$T\$O\$G\$K\$S	226
Db	2178	T\$A\$S\$Y\$T\$N\$E\$A\$T\$W\$P\$S\$A\$R\$L\$H\$O\$G\$R\$N\$A\$R\$P\$V\$N\$P\$K\$E\$M\$O\$V\$O\$F\$O\$T\$M\$K\$V\$T\$S\$V\$T\$O\$G\$K\$S	223
Qy	2270	L\$T\$S\$M\$T\$V\$E\$F\$L\$S\$S\$O\$G\$H\$O\$M\$T\$P\$O\$N\$G\$V\$Y\$K\$P\$O\$G\$N\$D\$S\$T\$P\$V\$V\$N\$S\$D\$P\$L\$T\$R\$V\$L\$R\$H\$P	2323
Db	2238	L\$T\$S\$M\$T\$V\$E\$F\$L\$S\$S\$O\$G\$H\$O\$M\$T\$P\$O\$N\$G\$V\$Y\$K\$P\$O\$G\$N\$D\$S\$T\$P\$V\$V\$N\$S\$D\$P\$L\$T\$R\$V\$L\$R\$H\$P	229
Qy	2330	Q\$S\$W\$H\$O\$A\$L\$R\$E\$M\$E\$V\$L\$G\$E\$N\$O\$D\$Y	2351
Db	2298	Q\$W	

RESULT 3


```

Db      1579 QASRPFYSSLSISYPDDQEOGAEPRHNFVOPNBTFTYFAKVOHNAAPTEDEFOCKAMAY 1638
      |||
Oy      1857 FSDVDLEKDVHSGILGILLVCHNTLNPANROTVQVEFLPTIDFKSKYFENNER 1916
      |||
Db      1639 FSDVDLEKDVHSGILGILLVCHNTLNPANROTVQVEFLPTIDFKSKYFENNER 1698
      |||
Oy      1917 NCRAPCNIMEDPFFKFNKRNFAHNGITIMDLPGLVNAOQRTRTLLSMGSENHSH 1976
      |||
Db      1699 NCRAPCNIMEDPFFKFNKRNFAHNGITIMDLPGLVNAOQRTRTLLSMGSENHSH 1758
      |||
Oy      1977 FSGHVFYRKKEEYKALYNLYPGVEFVEMLPKAGIMRVECLIGELHAGMSTLELYV 2036
      |||
Db      1759 FSGHVFYRKKEEYKALYNLYPGVEFVEMLPKAGIMRVECLIGELHAGMSTLELYV 1818
      |||
Oy      2037 SNKOTPLGMAHSHIRFOITASGOYGMAPKRLARLYSGSIANSTKEPFSMTKVDLLA 2096
      |||
Db      1819 SKECAPLGMAHSHIRFOITASGOYGMAPKRLARLYSGSIANSTKEPFSMTKVDLLA 1878
      |||
Oy      2097 PMHIGITGAGARKPSLIYISQFIINSLDGKKNQYRGNSTGLTAMFEGVDSGCIKH 2156
      |||
Db      1879 PMHIGITGAGARKPSLIYISQFIINSLDGKKNQYRGNSTGLTAMFEGVDSGCIKH 1938
      |||
Oy      2157 NIFNPPIIARYIRLPHYHSINSTLMELMCGDINSCEMPLGMSKRAISDAQITASSYFT 2216
      |||
Db      1939 NIFNPPIIARYIRLPHYHSINSTLMELMCGDINSCEMPLGMSKRAISDAQITASSYFT 1998
      |||
Oy      2217 NMFATSPSKARHLHOGSRNANPQVNNPKEMLOVDFOKTMKVTGYTQVSKSLTSMYV 2276
      |||
Db      1999 NMFATSPSKARHLHOGSRNANPQVNNPKEMLOVDFOKTMKVTGYTQVSKSLTSMYV 2058
      |||
Oy      2277 KEFLSSSDGCHQWTLFQNGKRYKPGNOSFPPVNSIDPPLLRKRYRHRQSRWQOI 2336
      |||
Db      2059 KEFLSSSDGCHQWTLFQNGKRYKPGNOSFPPVNSIDPPLLRKRYRHRQSRWQOI 2118
      |||
Oy      2337 ALRMEVLGCEADLY 2351
      |||
Db      2119 ALRMEVLGCEADLY 2133
      |||

RESULT 4
KPH05
coagulation factor V precursor [validated] - human
N.Alternate names: coagulation labile factor; proaccelerin
C.Species: Homo sapiens (man)
C.Date: 19-May-1989 #sequence_revision 02-Jun-1995 #text_change 08-Dec-2000
C.Accession: A56172; A42344; A28028; A27498; A23897
R.Citipe, L.D.; Moore, K.D.; Kane, W.H.
Biochemistry 31, 3777-3785, 1992
A.Title: Structure of the gene for human coagulation factor V.
A.Reference number: A42344; MUID:92232668; PMID:1567832
A.Molecule type: DNA
A.Residues: 1-2224 <CR>
A.Cross-references: GB:J05368
A.Accession: A42344
A.Molecule type: DNA
A.Residues: 48-58;79-89;120-130;191-201;239-249;313-323;368-378;428-437;461-471;533-542;
2070;2111-2120;2172-2181 <CR>
R.Jenny, R.J.; Pittman, D.D.; Toole, J.J.; Kriz, R.W.; Aldape, R.A.; Hewick, R.M.; Kaufm
Proc. Natl. Acad. Sci. U.S.A. 84, 4846-4850, 1987
A.Title: Complete cDNA and derived amino acid sequence of human factor V.
A.Reference number: A28028; MUID:87260886; PMID:3110773
A.Accession: A28028
A.Molecule type: mRNA
A.Residues: 1-857; 'R', 859-864; 'R', 866-924; 'E', 926-1763; 'I', 1765-2212; 'T', 2214-2224 <JEN>
A.Cross-references: GB:M16967
A.Note: parts of this sequence, including the amino end of the mature protein, were dete
R.Kane, W.H.; Ichinose, A.; Hagen, F.S.; Davle, E.W.
Biochemistry 26, 6308-6314, 1987
A.Title: Cloning of cDNAs coding for the heavy chain region and connecting region of hum
A.Reference number: A27498; MUID:88107560; PMID:2827731
A.Accession: A27498

```

```

A.Molecule type: mRNA
A.Residues: 1-1284; 'I', 1286-1600 <KAN>
A.Cross-references: GB:M17785
A.Note: parts of this sequence were determined by protein sequencing
R.Kane, W.H.; Davle, E.W.
Proc. Natl. Acad. Sci. U.S.A. 83, 6800-6804, 1986
A.Title: Cloning of a cDNA coding for human factor V, a blood coagulation factor hom
A.Reference number: A25897; MUID:86513665; PMID:3092220
A.Accession: A25897
A.Molecule type: mRNA
A.Residues: 1188-1215, 1315-2224 <KA>
A.Cross-references: GB:M14335
A.Note: parts of this sequence were determined by protein sequencing
R.Keller, F.G.; Ortel, T.L.; Quinn-Alten, M.A.; Kane, W.H.
Biochemistry 34, 4118-4124, 1995
A.Title: Thrombin-catalyzed activation of recombinant human factor V.
A.Reference number: A56139; MUID:95210278; PMID:7696276
A.Contents: annotation; thrombin cleavage sites
C.Comment: Factor V is activated by thrombin and partially by coagulation factor Xa
C.Genetics:
A.Gene: GDB:A5
A.Cross-references: GDB:119896; OMIM:227400
A.Map position: 1q23-1q23
A.Introns: 53/2; 84/1; 125/1; 196/1; 244/1; 318/1; 373/2; 432/3; 466/1; 537/3; 588/1
C.Function:
A.Description: acts as a cofactor, with calcium and phospholipid, for the factor Xa
A.Pathway: blood coagulation
C.Superfamily: coagulation factor V; discoidin I amino-terminal homology; ferroxidase
C.Keywords: blood coagulation; duplication; glycoprotein; phospholipid binding; plate
F:1-28/Domain: signal sequence #status predicted <SIG>
F:32-2224/Product: coagulation factor V #status predicted <MAT>
F:32-737/Product: coagulation factor Va heavy chain #status experimental <VAM>
F:32-345/Domain: A1 <DA1>
F:33-339/Domain: ferroxidase repeat homology <FO1>
F:346-691/Domain: A2 <DA2>
F:351-684/Domain: ferroxidase repeat homology <FO2>
F:692-1573/Domain: B <DOB>
F:1183-1461/Region: 9-residue repeats (Q-X-T/N-L-S-P-D-L-S)
F:1574-2224/Product: coagulation factor Va light chain #status experimental <VAL>
F:1574-1905/Domain: A3 <DA3>
F:1581-1905/Domain: ferroxidase repeat homology <FO3>
F:1667-1765/Region: phospholipid binding #status predicted
F:1906-2064/Domain: C1 <DC1>
F:1906-2064/Domain: discoidin I amino-terminal homology <DN1>
F:2065-2224/Domain: C2 <DC2>
F:2065-2224/Domain: discoidin I amino-terminal homology <DN2>
F:321-523/297,460,468,554,741,752,760,776,782,821,938,977,1074,1083,1103,1106,147
F:167-193;248-329,500-526,603-684,1125-1751,1907-2061,2066-2221/Disulfide bonds: #S
F:334-335/Cleavage site: Arg-Asn (protein C) #status predicted
F:363-693/1546/Binding site: sulfate (tyr) (covalent) #status predicted
F:376-377/Cleavage site: Arg-Ser (coagulation factor Xa) #status predicted
F:382-1338/Binding site: carboxylate (Asn) (covalent) #status absent
F:534-535/Cleavage site: Arg-Gly (protein C) #status predicted
F:737-738/Cleavage site: Arg-Ser (coagulation factor Xa, thrombin) #status experime
F:1046-1047/Cleavage site: Arg-Thr (coagulation factor Xa, thrombin) #status exper
F:1573-1574/Cleavage site: Arg-Ser (thrombin) #status experimental

Query Match 22.7%; Score 2617.5; DB 1; Length 2224;
Best Local Similarity 30.5%; Fred. No. 8e-139
Matches 768; Conservative 388; Mismatches 832; Indels 507; Gaps 77;
22 RRYVIGAVEISWDYQSDGLGELPVDARPPRPKSPFPNTSVV-YKKTLFVEFDHLENT 80
Db      32 ROFYVAAGISMSYRPE-----PNSLNLNLSVSEFKIYREYEPY-FKK 75
Oy      81 AARPRPMGLGPIPTIOEDYDVITLKNMASHVSHAAVGYMKASGAEVDQDSOR 140
Db      76 KRQSTISGLGPIVLAHVGDIDIKYHFNKADKRLSHHGGRIGRYSLEBSGYLDHFRPA 135
Oy      141 EREDKATPGGSRITTYQVLYKENGEMASDPLCLITSLSHVDLYKDLNSGLGLALYCRE 200
Db      136 EKMDAVALPGRHYEYEMSSDSGPHDDPCLNTHYSHENLIDFNLSGLGPLCKK 195

```


Db 2053 TLRLEQCEVNGCSPILGEMNGKIENKQITASSFKSMWGDYWEFFARLMAQGRVNAW 2112
 Qy 2239 RPOVNPKEMLQVDFQTKVTVGTGQVKSILSMYVEKFLISSODQHOMTFPONGK 2298
 Db 2113 QAKANNKQWLEIDLKIKITAITQGCFSLSSEMYVKSYTHSEQGEWKPYRLKSS 2172
 Qy 2299 V--RYFGNODSFYPVVNSLDPELITRYLRTHPOSWHOJALRMETLCEMODIX 2351
 Db 2173 MDKLTGSGNTYKGVAKNFPNPIISRTKIVPTMNOSTALMLKLGCG--DIT 2224

RESULT 5

742764
 coagulation factor V - mouse
 C:Species: Mus musculus (house mouse)
 C:Date: 11-Jan-2000 #sequence_revisions 11-Jan-2000 #text_change 09-Jun-2000
 C:Accession: T42764
 R:Yang, T.L.; Gul, J.; Behnmtulla, A.; Yang, A.; Mousaelli, M.; Kaufman, R.J.; Ginsburg, Blood 91, 4593-4599, 1998
 A:Title: The structure and function of murine factor V and its inactivation by protein C
 A:Reference number: 422270; MUID:98282202; PMID:9616155
 A:Accession: T42764
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-2183 <YAN>
 A:Cross-references: EMBL:052925; NID:g3219690; PID:g3219691; PIDN:AMC99553.1
 C:Function:
 A:Pathway: blood coagulation
 C:Superfamily: coagulation factor V; discoidin I amino-terminal homology; ferroxidase re
 C:Keywords: blood coagulation; duplication; glycoprotein; phospholipid binding; plasma: F;350-682/Domain: ferroxidase repeat homology <FOX2>
 F;1541-1864/Domain: ferroxidase repeat homology <FOX2>

Query Match 22.38; Score 2768.5; Db 2; Length 2183;
 Best Local Similarity 29.88; Pred. No. 2,9e-136;
 Matches 749; Conservative 408; Mismatches 852; Indels 503; Gaps 73;

Qy 8 CFFLCIL---RFC-----FSATRRYYLAGVELSDYQMSDGLGELPYDAFPPRPKRS 56
 Db 7 CFFLLVYLGTRMAGWGSQAEAAQLNQFYVAAGILNNYHPE-----PTD 51
 Qy 57 PPFNTSVYKKTLFVEPTDHLFNIAKPRPMMWGLLPTIOAEYDPTVITLKMASSHPVS 116
 Db 52 PSLNSIPSFKTIYRREY-EQYFKKEKPRSSNSGLLGPLVLAEGVDIKAFHFNKDKRLS 110
 Qy 117 LNAVGVSTWKASGEGAEYDQTSOREKEDKYPGGSHTYWQYLKENGSMASDPLCLTYS 176
 Db 111 IHPGKIKYSKFSGASADHTFPAERKDAVAPEEYTEMIVSEDSGPTPDDPCLTHI 170
 Qy 177 YLSHYDLVKDLSGLIGALLVCREGSLAKEXTQTL--HKFILLFAVFDGKSMHSETKNS 234
 Db 171 YSYENLQDPNSGLIGPLILCKKGLTFEDGQKMFQOHVLLFAVFDSSKS----- 222
 Qy 235 LMQDDNASARAMPKMTVNGVYNSRLPGLIGCRKSYWHVIGMTGTEPVISIFLEGT 294
 Db 223 -----RSQSPSLMTTTFYFNKIMPDITVCAHDVSHMLIGKSSGELSSIFNQY 274
 Qy 295 FLVRNHRQASLEISPTITLTAOTLLMDQGLFLFCHLSSHQDGMGAAYKVCSEEPQL 354
 Db 275 LBNQHKSTVTLVSAISTANNTMSPGKRWVSSLLPKHYQAGMOAYDINDCKKTR- 333
 Qy 355 RKMNEAEADYDDDLTSEMDVYRFDDNSPFIQIRSVAKKHPKWTWYVIAAEEEDMDY 414
 Db 334 -----SPK--TILREQRRYKMKWEYFIAAEVIMNY 362
 Qy 415 APLVLAAPPDRSKSQYLNQPGRIQKTKYKFMAYTDETF--KTRDAIQHESGLLGL 472
 Db 363 APVIRPANDKIKYRQHDNFSNQIGKHKKYKYYRQYEETFTKTRNPISIKSGILGVY 422
 Qy 473 YGEVGDILLIIFKKNQASRPYNIYPHGIDVRYLSRLPKGV-----KHLKDPILPG 525
 Db 423 RAQVVDTLKIVFKMAASRPYSITPHGV-----FSPYEDGINSSTSGSHTTINPVG 476

Qy 526 EIKFKYTVTEVDGPKSDPRCLIRYSSFPVNNERDLASGLIGPLILCYKESVDQNGNI 585
 Db 477 EFTFYKNNILFEDEPEENDAOCTRPYSDVYVRJAGLGLLICKSSLDQORVOR 536
 Qy 586 MDDRKNYILSFEDERSWYLTENTQRLPMPAGVQLEDEPEQASINMSHNGYVSDIQ 645
 Db 537 VADIEQAQVAVYADEKMSVIEDNINKCENPBEVRRDPPKYEENINSHINGVPESSIS 596
 Qy 646 -LSCLEHVAWYILSIGAQDPLSVFSGCYTRKHKMYEDLTLFPFSGSYVMKMEAP 704
 Db 597 TLGCFDTPVQMHFCVGHDDILTHFTGSHFTGRHDDTLTPMRGSAVTVMDNV 656
 Qy 705 GLMILGCHNSDFERNKMTALLKVCSSCDKNTGDY-YEDSYE----- 743
 Db 657 GTWMLTMMNSNPRRMILRFRFDVACNR--DYDNEDSEYIEPEAPYSMTTRRHISLE 713
 Qy 744 -----DISAVLLSKNNALTEPFSFQNSRHPSPROQAFATIP-ENDEKTDWFA 793
 Db 714 NEFGIDNEDDQYLLASSLGI--RSFKNSLNP--EENDFNLJALENSESIFSP--- 766
 Qy 794 HRTPRKIQWVSSDILMLLROSPTPHGLSLDLOEKAYETSDSPSCA----- 843
 Db 767 ---STDRKVDSSNRILSKTIYN-----NLKDDQRL-----PGSGATVYAGTLRL 810
 Qy 844 --IDSNNLSMTFHRPOLHSGDNVTFPESGQLRLNEKLGTTAATELKLDPKVSIS 901
 Db 811 IGLDENFVLSNSTERSSSYHENDME-NPOSNITMYLLPLGPKG----- 854
 Qy 902 NMLISTPBDMLAGTDNTSSLSGPPSMFVHDSOLDVTL--FGKSSPLTESGGPLSLSE 959
 Db 855 -----SGNRBDQPKTITKGRHMKKHFMSMAKPAKGTGRBSNPKNSYSG--MKSE 904
 Qy 960 ENNDSKL-----LEGSLN-----SOSSW--GNVSSRESRLTKGRARAGPALTF 1004
 Db 905 EDIPSELPLKQKITSKFLARKMRVASSGSEYELIAANGSDIDVDKL----- 951
 Qy 1005 KDNALFKVYSISLTKTKTSNNSATNRKTHIDGPSLLIENSPYWNILSDTERKVTPL 1064
 Db 952 -TNSPQONITVPRGSESTHTYTKRPDLTFPGVGHKSPHVRD--EENSQOKQLF 1008
 Qy 1065 IHDRLMDKNATALRLHMSNKTSSKNMEVQOKKGPTRPDAQPMQSFKKMLFLPES 1124
 Db 1009 ITR-----KKKNKKLAL-----HSPLSRPGDP----- 1033
 Qy 1125 ARVIGTRGKNSLNSGQSPKQVLYSLDPEKSYVGQNFLEKAKVYVKGFTYDVLKE 1184
 Db 1034 -----LRGHNHP----- 1041
 Qy 1185 MYFPSSRNILFTNLNDLHENTHNOEKRIQIEIEKKEELLQENYVLPQI--HYTGKNF 1242
 Db 1042 --FPDRRLNLSIL--LHKS--ETALSPLDNLQTSMSSTRSLPQVNOYSKNDYEQM 1093
 Qy 1243 MKNLFLLSTRONGVSGYDAVAVYLDQFRLSNDSTNRKTHAFKKGEEBNL----- 1296
 Db 1094 SSLSDLX--QSVAAEHSHPTPA-QDDQDHSHTDPSYSSPDELSGLDYDLSHDFY 1149
 Qy 1297 EGIQNGQYIYKACTRISLPTSGQNFQBSKALKQFLPL--EETLEKRII--- 1351
 Db 1150 DDIG-----LTSFPQOQSKSPSSDDQALPSDSLTFPLSPBEDDTITPD 1197
 Qy 1352 VDDTSTQWSKMKHLPTSTLTQI---DYNEK-----BKAITQSPSLDCLFRSHSTPOAN 1403
 Db 1198 LDQLLSPEDNQTKTSPD-LGVPLSPDQKTSPPDLGQVSLSPDQKTS--PDLG 1253
 Qy 1404 RSPFLPI--AVSSFPISIRIYLT-----RYLFQONSRLPAASRYKKQSGVQESHLFG 1456
 Db 1254 QVPLSLDQNTTIS-PDLGVPLSPDQNTTISPDQLQVPLSSDQKTSPPDLGVPLFP 1312
 Qy 1457 AKKNNLSLAILTEMTQDQREVSLSATNSVYKAYEVTVLPKPD-----LPKTSGV 1511
 Db 1313 EDNQVFLDLSQVPLSDQNG-----ETSIDLTLSPDFQVYVLSPLDQLPLSDNSOV 1368
 Qy 1512 ELLPKVH-----LYQKDLPTLETNSGSGHLDVBSGLGCTGCAIKKWEANRP 1560

Db 1369 TVSPDLSLTLPSPDENETIIAPLDGQVTLSP-----DLIGTN----- 1405

Qy 1561 GKVPLRAVATSSATPEKLLDPLADNHYGQIQREKMEQSESPKFAKKKDIISL 1620

Db 1406 ---PALNHGKAKSSADPLQASYP--DSCQASLP--ELNRTLPHPDLHIPTPPSPPL 1458

Qy 1621 NAGESNHAIAINEGONKPELEYTMAKQGERLCSQNPV---LKRHREITRTTLOSD 1677

Db 1459 N-----NTSLSRKFPNPLVVLGSLRVDDDELVPSEE 1490

Qy 1678 QEEIDYDPTISYEMKEDFDIYDEENQSPSPKRTNR-----YFI 1719

Db 1491 PERIDED-----YAEDDFVYNDPRTDRTDVSSNNPDTIAAMVLRGSGHKFYFI 1544

Qy 1720 AYVERLMDYGSSSHVLRNMAQSGSPV---FKYVYQETDSDSTQPIRLRELNHG 1776

Db 1545 AAEETWNTAEFQASEM--DHEDGHPKDTYTKVFRKTKLDTSTFYSRDPRAVEEHLG 1602

Qy 1777 LAGPIRAVEDNINWFRNQASRPYSFVSSLSYE-----EDROGAEPKRNKVP 1828

Db 1603 ILGPIRAVEDNIVYRKRNLASRPYSLSHAGLSYEKSEKRTYEDSESPFQEDDAVP 1662

Qy 1829 NEFTYFMVQOHMAPTDEDFCSKAAAFESDYDLKDVHSGILGPILVCTNTLNPAHR 1888

Db 1663 NSSYTYVNHATKRSQEPENPGSACRAWAYSAVNVVERDIHSGILGPILVCTNTLNPAHR 1722

Qy 1889 QVTVQEFALFTIIPDETSMYFETENMERNCAPCNLOMEDPPEKENTRPAINGYIMTL 1948

Db 1723 PMDRREFVLFWFDEKSMYTEES--KGR-----RIESEPKNAKFTALNGMITYN-L 1774

Qy 1949 PGLVMAODRIRWYLLSGNSNENHSHSGHVFYVYRKKEEKMAALNLYGVFEVEKL 2008

Db 1775 PGLMYEQEWVRHLNLNNGSSRDIVHVFHQTLDNRTKHQGVNPLPLGSGFKLEK 1834

Qy 2009 PSKAGIRVECLIGHSLHAGMSTLELVYSNKCOTPLGMAASHIRDPQITASGOYGMAR 2068

Db 1835 ASKRGWMLDTEFGENQVAGMQPFLIIDEKCEMPMGSLSTGVISDQIKKASELITWERN 1894

Qy 2069 LARLHVSQINAMSTKE-----PPS-WIKYDLAPLIIHGKIQGAAOKSSLYISQFI 2122

Db 1895 LARLNAGSVMASIEKRALDPLIKPIAYQDMKEVYVGIQOGAAHLYLAKSCPTTEFPV 1954

Qy 2123 MYSLDGKQKQYRKGSTGTLMVFEFGANDSGIKAHNIFNPPIIARIYLRHPHYSISSTR 2182

Db 1955 AYSSDOTNMQIFRGKSGSVMYFTGNSDSTIKENLDPLVARYIRIHPTKSYNRPPLR 2014

Qy 2183 MELMGCDLNSCWPLGMSKASISDAQITASSYFTNMFAF--MSPSKARLHIOGRSNAMRQ 2241

Db 2015 LELQGEVNGCSPTLGLDEGRIDQKITASSFKSWMGDYEPESLARLNAGRVNMAQK 2074

Qy 2242 VNNPKEMLDVQOKMKYTGVTGVSLLTSMYKEFLISSQDGHOMTLFQONKRV-- 2299

Db 2075 ANNNKQMLQVDLIKIKRTATVTOGCSLSSEMYKVSISIOYSDQYAMKPYRKSSWD 2134

Qy 2300 KVFQGNDSFTYVNSLDPPLLYNLIHPOSWHOIALRMEVIGCEADLY 2351

Db 2135 KIFEGNSNTRKGHKNEFNPPIISRFIRIIRKTNOSIALRELFGC---DIY 2183

RESULT 6

KP005

coagulation factor V precursor - bovine

C:Species: Bos primigenius taurus (cattle)

C:Date: 04-Mar-1993 #sequence_revision 28-Apr-1995 #text_change 11-Jun-1999

C:Accession: A42580; A36497

R:Quinto, E.R.; Remon, C.T.; Mann, K.G.; MacGillivray, R.T.

J. Biol. Chem. 267, 2971-2978, 1992

A:Title: The complete cDNA sequence of bovine coagulation factor V.

A:Reference number: A42580; MUID:92147638; PMID:1737753

A:Accession: A42580

A:Molecule type: mRNA

A:Residues: 1-2211 <GUI>

A:Cross-references: GB:M81440; NID:g163037; PIDN:AAA0512.1; PID:g163038

A>Note: sequence extracted from NCBI backbone (NCBIIN:80774, NCBI:P.80776)

R:Kafatis, M.; Jenny, R.J.; Mann, K.G.

J. Biol. Chem. 265, 21580-21589, 1990

A:Title: Identification and characterization of a phospholipid-binding site of b

A:Reference number: A36497; MUID:91072354; PMID:2234316

A:Accession: A36497

A:Molecule type: protein

A:Residues: 1566-1570, 'X', 1572-1581, 'X', 1583-1584, 1673-1676, 'X', 1678-1679, 'X', 16

R:Xue, J.; Kafatis, M.; Silveira, J.R.; Kung, C.; Mann, K.G.

Biochem. J. 33, 13109-13116, 1994

A:Title: Determination of the disulfide bridges in factor Va heavy chain.

A:Reference number: A5979; MUID:95034740; PMID:7947716

A:Contents: annotation

A>Note: 566-Cys and 617-Cys were shown to have free sulfhydryls

C:Comment: Factor V is activated by thrombin and partially by coagulation factor

C:Function:

A:Description: acts as a cofactor, with calcium and phospholipid, for the factor

A:Pathway: blood coagulation

C:Superfamily: coagulation factor V; discoidin I amino-terminal homology; ferrox

C:Keywords: blood coagulation; duplication; glycoprotein; phospholipid binding;

F:1-28/Domain: signal sequence #status predicted <SIG>

F:29-221/Domain: signal sequence #status predicted <SIG>

F:29-221/Product: coagulation factor V #status predicted <MAT>

F:29-741/Product: coagulation factor Va heavy chain #status predicted <MAT>

F:29-345/Domain: A1 <DA1>

F:33-329/Domain: ferroxidase repeat homology <FO1>

F:346-695/Domain: A2 <DA2>

F:351-688/Domain: ferroxidase repeat homology <FO2>

F:696-1564/Domain: B <DOB>

F:1175-1537/Region: 9-residue repeats (Q-X-T/R-L-S-P-D-L-S)

F:1565-2211/Product: coagulation factor Va light chain #status predicted <VAL>

F:1565-1892/Domain: A3 <DA3>

F:1565-1892/Domain: ferroxidase repeat homology <FO3>

F:1654-1752/Region: phospholipid binding #status predicted

F:1893-2051/Domain: C1 <DC1>

F:1893-2048/Domain: discoidin I amino-terminal homology <DN1>

F:2052-2211/Domain: C2 <DC2>

F:2052-2208/Domain: discoidin I amino-terminal homology <DN2>

F:1657-1937-248-329-499-525/Duplicate bonds: #status experimental

F:225-229-297-382-460-553-587-756-774-780-902-952-964-1044-1053-1062-1071-107

F:324-335/Cleavage site: Arg-Asn (protein C) #status predicted

F:363-697-1537/Binding site: sulfate (Tyr) (covalent) #status predicted

F:376-377/Cleavage site: Arg-Ser (coagulation factor Xa) #status predicted

F:533-534/Cleavage site: Arg-Gly (protein C) #status predicted

F:607-688-1712-1738-1894-2048-2053-2208/Duplicate bonds: #status predicted

F:741-742/Cleavage site: Arg-Ser (coagulation factor Xa, thrombin) #status predic

F:1034-1035/Cleavage site: Arg-Ser (coagulation factor Xa, thrombin) #status pred

F:1564-1565/Cleavage site: Arg-Ser (thrombin) #status experimental

Query Match 22.1%; Score 2744; DB 1; Length 2211;

Best Local Similarity 30.2%; Pred. No. 5-6e-135;

Matches 765; Conservative 381; Mismatches 834; Indels 550; Gaps 80;

Qy 22 RRYLGAVELSDYQSDLGELPDAFPPRPKSPFNISYVYKTLVEYFIDHLFNIA 81

Db 32 RQFLVAAGSIRNNYR-----PESTHL-----SSNPEST--FKYIVREY-EAVFQKE 76

Qy 82 KPRPMGLIGPTIQAEDYDVYITLKNASHPVSLHAGVSYKASSEGAEYDDQTSORE 141

Db 77 KQPSRTSGILGPTLAEVGDIMKHEFKNAKRPILSHAGIKYSKFSQASYSDDHTLPKE 136

Qy 142 KEDDKYFGSRTYVQVYLKENGPMASDPICLTYSYLYSHVDLVNDLSGLIGALVCREG 201

Db 137 KQDVAAPQDEYTYEMISHSRGTTHDDPPCLHITYSVNLVDFRSGILGPIIDCKKG 196

Qy 202 SLAKERTQTL--HKFLLFVFPDGSKMSHSEIKNSLMDQDRAASARAPKMTYVGYNR 259

Db 197 TVLEDGTVQFKQVNLFAVFDKSNKQTS-----LMTYVNGYVNG 240

Qy 260 SLQGLICRKSQVYHVGCTQPEVHSLFGLGHFLVNRHQALESIPITFLAQTL 319

Db 241 TMDPIVCAHDHISWHLIGMSSGPELFSIHFNQVLEQNNHRIKISATIVASSTYAMTV 300

Qy	320	MDLOGLFLPHSHISHOHOMEMAYKVADVSCSEEBZOLMKNINNEADYDDDLTJSEMDYVFR	379
Db	301	SPERWTIATLIPRHQAGMOAYIDINKCAKTRNPK-----LTBQ-----	343
Qy	380	DDDSBSPFIOIRSVAKKHQWYVYIAAEEDMDAPYAPADDSRYKSOYLINBQIG	439
Db	344	-----RKHKKRETFIAHEEYIMDAPILPAMDKKRSJLHLDNBSNRIG	388
Qy	440	RKXKVFMAVYDEFTKR--EAIQESGLPGLXGEGDTLLIFKNOASFPINYPH	497
Db	389	RHKYKVVYKQYODSEFTKLEDPESSEDGILGPIIAQORPLTIVERNASHSYSTPH	448
Qy	498	GIT-----DVRLYSRRLPKGVNHLDPFLIPGLGFYKKTWYVEGPKSPDROL	548
Db	449	GVTSPDYDNENSSSTSGSNMTRAVR-----PGELYTKNNILSEPEENDAOCL	500
Qy	549	PYYTSSPYVNERDPLAGLIPGLICYSKVODRGNOIMSDKNITLPSYDEKRSYLTLE	608
Db	501	TPPYTSNVMDITRLDLAGLIGLILCKRSJDRGIGRADIEQOAVPAVDEAKSMYIED	560
Qy	609	NIQRLTPRPAQVODLPEREQASNTM-----SINGYVDSLO-LSVCLHEAVYILSIG	662
Db	561	NIYKCEPKEPVKADPKPTESNLSNFTLPALINGVPESTPLIGFCFDYVOMHCSVG	620
Qy	663	ACQDFLVSFFSGYTFKHKKVYEDPLTPPESGSETVMSMENEGMLILGCHNSDFNRKGT	722
Db	621	TQNDILTIHFTGHSFYGRHEDYLTLEPMQGESVTVMDVNGWMTLTMNSNPKSLR	680
Qy	723	ALLKVSQCDNNTDXYEDYEDY-----SAYILSKNNA-----	755
Db	681	LEFRPACIENRDD--DSEYELIYEPGSGTAMTKKIHSSSEIENDADSDYODELALI	737
Qy	756	IEPRSEFSONSRHSTROKOFNATPIP-ENIETKTDWFMHRTPMPKIONVSSDMLMLR	814
Db	738	LGLRSEFRNSSI--NQKDELTALALEKDESETPP-SANRS-----LDNSSS-----	783
Qy	815	QSPRHGISLSDLOEKAYTFSDDPSCA-----IDSNLSISEMTHPRQOLHSG-----	864
Db	784	RSHVSRLLAKNPFASJKTLLHLEAPAGSPLEHAGDKSNALNP-----PMASHSPSYE	838
Qy	865	----DWTFPESGLQ-----PLNEKLGTTAMBLKLDLFKYSYSTNNL	904
Db	839	DREHRIPLSDTVETSLIPRGVGFKNRPAKHQRPYGVGRQAARKHFSQTRFPAKTRFL	898
Qy	905	ISTIPSDNLACDTNNS--LQPPSMVHYDSOLDITLFGKSS-----PLTSGGP	954
Db	899	-----SODNSSSSRMGP-----WEDISDILLQOQDPFKILNEMELVSEKS	942
Qy	955	LJLSEENNDSKLLSEGLMSOES--SMGKNVSTESGHLFGKRAQAPLLH-----	1004
Db	943	YELLIDANENKTYNK-LPNSPQNDSSRTWGENIFPKNS--HGQSGHPFLIVTRRKLQ	997
Qy	1005	----KDAKPRVYSISLTKNTKSNSKATNKTTHIDPSPILJINBSYWO-----NILES	1054
Db	998	DNQDRNRSRLKEGPIPLR-----TRKKKEEKPAYHNPALSRSHPLKEVNASFS	1048
Qy	1055	DTEFAKVTPLIHDMLMDKNATLALNEMSNKTTSSKNEMVQOKKEGPIPPDAQNPDS	1114
Db	1049	DREHN-----HSLTLHASNETSLSD--DNQTFPMNLSLA-----ASLDP--	1087
Qy	1115	PFKMLFIPESARMLQRTQKNSLN--SGQGPSPKOLV-SLGEKSEVGONFLSEKKRVV	1171
Db	1088	-----HDQSPNDTSSQSSPPDLYPVSPSEEYQ-----	1117
Qy	1172	GKGFTKQVGLKWEVFPSSRNLFJLNDNLHNNTHNOKKLOEIEKKEFLIQEYVLP	1231
Db	1118	-----IIP-----IQDSDPHSTPASHRSDD	1139
Qy	1232	QHHTVYTGKNNKMLLSTRONVGSGTD--GAYAPYDQFNSL-----	1273
Db	1140	PTHSTAPSNRP--PTQSOQIPNDILDNRLAPTDVSOQLPFLLEWYQATSYLSDS	1194
Qy	1274	NDSTNRKKHHAHSAKSGEENLJEGLOMQKQIVERTYACTTRISBNTSQONFVQSKRA	1333

Dd	1195	QPSISPDLOGMSPDPGQESLSPDLQ-----	-TSLSPDLQESLSPDLQGT	1242	
Qy	1334	LKQPRFLPLEET---ELEKTIYDQTSQWMSKNNKHLTP-----	-SPLTDIOVNEKEXGALTO	1386	
Dd	1243	LSPP--PQSLSLSPDLGQALSPDS-----	-QESLSPDLQGTALSPDPQESLSPDLQO	1294	
Qy	1387	SPLSDCLTR---SHSIPQANSPLPIAKVSPFSIRPIYLRIYLFQONSHLPASARYK	1443		
Dd	1295	TLSLSPDLQSSLSLSPDLGQALSPDP-----	-QSLSLSPDLGQALSPDP-----	PSQESLSP 1344	
Qy	1444	DSGVQESSHFLQAKKNNLSLALITLMTGDDQREYVSGLSGTSATNSVYTKKVENTYLPKPD	1503		
Dd	1345	DLGQTSLSPLD--GQESLSPDLGQALSPDSSQSRSL-----	-SPD	1381	
Qy	1504	LEPKTS-----GKTELLPKVHYIYQKDLPEPTSSNGSPGHDLVQESLLOGTGEGAIKWNAN	1558		
Dd	1382	LQGTSLSPDLQGSLSLSP-----	-DLQGTALSP-----	DLQGSLSLSPDL-----	1418
Qy	1559	RKGKAPFLRATATSSAKKTPSKLDDPIAMDMHYTOIPIKXEMSOEKSEPTAKKKDPT-	1617		
Dd	1419	--GQPL-----SPDLSLSL-----	-SPDLSQDLQKQSP	1446	
Qy	1618	-LSLN-----ACESNHAIAAINEGONPRLEIYTAQGTERRLCQO-----	-NPVYLRH	1655	
Dd	1447	PLDINQTSHTSSQSSLSLPRFGQTFPMADIGMPSPPDSTLNTNTEPFEFNPVYVGL	1506		
Qy	1666	QRE-----ITFTTLOSQDEE-----	-IDYDQTSIVENKKEDPDIDYDEBQNSP-----	1707	
Dd	1507	SHDDDDYELIIPKQKESSEEDYGEFEFVAYNDPYOTDLRT-----	-DINSRNPDLNIA	1560	
Qy	1708	--RSFOKTRHFHFAIAVRLMDYQMSSPHVLRNAQSGS--	-VPO--FKRYVQOFT	1758	
Dd	1561	WYLRASNTGNRRYYIIAAEISIMDYS-----	-KFGQDDVDYDEPQYKRYVRYKL	1611	
Qy	1759	DGSFPOPLYRGEIENHLGLSPYIAAEYINWAFPRNQSAPVSPFSLISVE-----	1812		
Dd	1612	DSTFLKLDPOQEYBEHHLGILGYIAEYDVIQYRKKNLASPPFSLAHLSSTREKSSEK	1671		
Qy	1813	--EDORQAEPRKNFYKPEMETKTYPWKYQHMAAPTEDEPQCAANATYSVDLEKDYHSL	1870		
Dd	1672	TYEDSPSEWEKEDALAIOPKTTYTYWHAATRSBGENPSSACAAWATYSAVNPBKDIHSL	1721		
Qy	1871	IGPRLVCHTNTLPNABGRQVYQOEPALEFTTIDETKSWYFENNERCRAPCIOMEDPT	1930		
Dd	1732	IGPRLIICKGLDKETENMVDUREFVLLFVWFDEKKSMTYDKPKPRBSRRAS-----	1785		
Qy	1931	PEENTPRFAINGYIMTLPGLVMAODQIRHYLLTSMGSENHLSHESGVAFYVRKKEEY	1990		
Dd	1786	VANSHEPAINGMUYN-LPGLMRYBQENVRHLHLNNGSROIHVHFHGGTLENGTOOH	1844		
Qy	1991	KMALNLYRGVFEVYEMLSKAGIRVBCILGHLHLAGSTLFLYNSKOTPLQAMASH	2050		
Dd	1845	QGVMPPLPFGSEFKLEKKASKGMMHLLDTEGELQAGQTFPLVLDRECKMFWGSLTOL	1904		
Qy	2051	HDQITNAGSGOYGMAPKILARLHYSGSINAW-----	-STK--EPFWSIKVLDLAPMIIHGK	2104	
Dd	1905	IASDQIOASEWGMPEKCTKLARLNNGGSINAMIAIEKLSITEFNEPPIQYDQKKEVLLTGIO	1964		
Qy	2105	TGQAGQKQESSLSYISOFTINYSLDGKMKQYTRNGSSTQTLWVFGGANVDSGSKRNIENPPI	2164		
Dd	1965	TGQAHYILKPYTTEFCVAYSLDRKNMFIIEFGNSSTNWAYFQGSNDSASTIREKNOIDPPVY	2024		
Qy	2165	AYYIKLPHTHYSIRSTRRLMELGDCOLNCSMPJGMSKSIASQATASSTYFMFAT--	VS	2223	
Dd	2025	AYYIIRISTYSGKPPALRIELQGCVCVNCSPJGMSKGIENKQITASSFRKSKMWNMYE	2084		
Qy	2224	PKAKLHLQGSNAAMRQVNNKEXLQYDQTKMYKTGYTQOYKXSLTSMYKKEFLIS	2283		
Dd	2085	PLALATLNQGVNNAWQAKNNNNNOQLADLKLTKKITALYVQGCSSLSSEAYKSTHYH	2144		
Qy	2284	SODGHQWTLFEQNGKY--KVQGNQDSFTTPVNSLDPPLTRYLRIRHPOSMVHOIALRME	2341		

Db 2145 SDOGTDMKPYREKSSMYDKIFEQNNVNGVHNFNPILSRIFRIPIKTNWSTALRLE 2204

QY 2342 VTCGEADOLY 2351

Db 2205 LFGC---DMT 2211

RESULT 7

A25945

coagulation factor VIII - pig (fragment)

C:Species: Sus scrofa domestica (domestic pig)

C:Date: 16-Aug-1988 #sequence_revision 16-Aug-1988 #text_change 21-Jul-2000

C:Accession: A25945

R:Tool: J.J. Pittman, D.D. Orr, E.C. Murtha, P. Wasley, L.C.; Kaufman, R.J.

A:Title: A large region (approx. 195 kbp) of human factor VIII is dispensable for in vitro

A:Reference number: A25945; MUID:86287369; PMID:3016730

A:Accession: A25945

A:Status: nucleic acid sequence not shown; not compared with conceptual translation

A:Molecule type: DNA

A:Residues: 1-869 <700>

C:Superfamily: coagulation factor VIII; discolidin I amino-terminal homology; ferroxidase

Query Match 17.6%; Score 2190; Db 2; Length 869;

Best Local Similarity 45.2%; Pred. No. 1.2e-106;

Matches 493; Conservative 122; Mismatches 251; Indels 224; Gaps 15;

QY 705 GIMWICGHSDFRNGMTALKVSSCDKNTGYEDYEDISATLTKNNATPRSFSON 764

Db 1 GIMWICGHSDFRNGMTALKVSSCDKNTGYEDYEDISATLTKNNATPRSFSON 60

QY 765 SRHSTRKOKFNATIPNDIEKTDPMFARHPMKRIONVSSDILMLRQSTPHGSL 824

Db 61 SRHSTRKOKFNATIPNDIEKTDPMFARHPMKRIONVSSDILMLRQSTPHGSL 119

QY 825 SDLEAKYETFDSDSPGALDSNNLSSEMTHERPOLHSGDMVTPPSGLQLTKNEKLT 884

Db 120 SDLEAKYETFDSDSPGALDSNNLSSEMTHERPOLHSGDMVTPPSGLQLTKNEKLT 167

QY 885 TAATELKIDPFVSTSNLIS--TIPSDMLAAGDNTSSGPPSMVHYDSOLDLTFLG 942

Db 168 TAATELKIDPFVSTSNLIS--TIPSDMLAAGDNTSSGPPSMVHYDSOLDLTFLG 223

QY 943 KKSPPLEBSGGLSLSENNDSKLSGLMNSOESMKNVSTESGRLEKGRANGPAL 1002

Db 224 KKSPPLEBSGGLSLSENNDSKLSGLMNSOESMKNVSTESGRLEKGRANGPAL 270

QY 1003 LTKDMALFVYSILTKTKTSNNSATNRKTHIDGSLIENSFVWONILSDTEKRYT 1062

Db 271 LTKDMALFVYSILTKTKTSNNSATNRKTHIDGSLIENSFVWONILSDTEKRYT 315

QY 1063 PLIHDMMDKNAFALRLHNSNKTSSKNMEWQOKKEGPIPPDAONPMSFFKMLFLP 1122

Db 316 PLIHDMMDKNAFALRLHNSNKTSSKNMEWQOKKEGPIPPDAONPMSFFKMLFLP 332

QY 1123 ESAMWIORTHGKNSNSGSPKOLVSLGPEKSVYEGONFLSEKNKYVKGKGTQDVL 1182

Db 333 ESAMWIORTHGKNSNSGSPKOLVSLGPEKSVYEGONFLSEKNKYVKGKGTQDVL 385

QY 1183 KEWYFSSRNPLEYIMDLHNNTHNOEKIOEIEKKEKTLIOENVLPOITHTVGTAKF 1242

Db 386 KEWYFSSRNPLEYIMDLHNNTHNOEKIOEIEKKEKTLIOENVLPOITHTVGTAKF 445

QY 1243 MKNLFTLSTROWESSYDGAAYALQDFRLSDNSTRKTHAHFSKKGEEELLEGANQ 1302

Db 446 MKNLFTLSTROWESSYDGAAYALQDFRLSDNSTRKTHAHFSKKGEEELLEGANQ 505

QY 1303 TKQIVKVACTWRISPNISQONFVTOKSRKALQKOPPLETELEKTIYDDSTQWMSKN 1362

Db 506 TKQIVKVACTWRISPNISQONFVTOKSRKALQKOPPLETELEKTIYDDSTQWMSKN 548

QY 1363 MKHLPTSTLTQIDYNEKKGALITQSPISDCLTRSHSIPQANRSPPLAKVSSPSTIRPY 1422

Db 549 -----

QY 1423 LTRVLEFQDNSSHLPAASYRRKSDGVOESSHFLQAKKNNLSLILTEMTQDQREVSLG 1482

Db 549 -----

QY 1483 TSATNSVYKRVKENTVLPKPDLPKTSKVELLPKVHYIQDLPFETNSGSPGILDLVEG 1542

Db 583 KSMAGFLASGKLEKAVLSAGLSEAGKEFLPKRVYRBDLLPQTSNVSCABDGLQGE 642

QY 1543 SLQGEKGIKMNENAPGVLPRLVATFESSAKPPSKTLDPGLAMDHYGTOLPKREKMSQ 1602

Db 643 IRLQKTRGVMLKMNVRP-----KRPSTLDP-----PMPK-EMESL 679

QY 1603 EKSPETAFKKDDIT-LSLWACSNHAIATYQNKRELETWMAQGTERTLSQNPV 1661

Db 680 EKSPETAFKKDDIT-LSLWACSNHAIATYQNKRELETWMAQGTERTLSQNPV 739

QY 1662 LKRRQREITRTTQSDQDEIDYDRTISVEMKKEPDYDDENQSPRSFQKTRHYFLAA 1721

Db 740 LKRRQREITRTTQSDQDEIDYDRTISVEMKKEPDYDDENQSPRSFQKTRHYFLAA 799

QY 1722 VERLMDYKSSSPHYLANRQSGSVFQFKVYFOEFTDGSFTQPLRYGELNHLGLGPPY 1781

Db 800 VERLMDYKSSSPHYLANRQSGSVFQFKVYFOEFTDGSFTQPLRYGELNHLGLGPPY 859

QY 1782 IRAVEEDNIM 1791

Db 860 IRAVEEDNIM 869

RESULT 8

KOHU

ferroxidase (EC 1.16.3.1) precursor [validated] - human

N:Alternate names: ceruloplasmin

N:Contains: ferroxidase long form (CP-1); ferroxidase short form (CP-2)

C:Species: Homo sapiens (man)

C:Date: 31-Aug-1980 #sequence_revision 12-May-1995 #text_change 08-Dec-2000

C:Accession: A25443; A24165; A35450; A00524; I59067

R:Koshikawa, M.L.; Funk, W.D.; van Oost, B.A.; MacGillivray, R.T.A.

A:Title: Complete cDNA sequence of human ceruloplasmin.

A:Reference number: A25443; MUID:8625737; PMID:2873574

A:Accession: A25443

A:Molecule type: mRNA

A:Residues: 1-1060; 1065-1069 <ROS>

A:Cross-references: GB:M1369; NID:q180255; PIDN:AA51976.1; PID:q180256

A:Note: this is the short or CP-2 alternatively spliced form

R:Mercer, J.F.B.; Grimes, A.

FEBS Lett. 203, 185-190, 1986

A:Title: Isolation of a human ceruloplasmin cDNA clone that includes the N-termin

A:Reference number: A24165; MUID:86275241; PMID:3755405

A:Accession: A24165

A:Molecule type: mRNA

A:Residues: 1-40; 549-599; 784-829; 919-952 <MB>

R:Yang, F.; Friedlrichs, W.E.; Cupples, R.L.; Bonifacio, M.J.; Sanford, J.A.; Kork

J. Biol. Chem. 265, 10780-10785, 1990

A:Title: Human ceruloplasmin, tissue-specific expression of transcripts produced by

A:Reference number: A35450; MUID:90285218; PMID:2335023

A:Accession: A35450

A:Molecule type: DNA

A:Residues: 1007-1064 <YAN>

A:Cross-references: GB:J05506

A:Note: this is the long or CP-1 alternatively spliced form

R:Takahashi, N.; Orel, T.L.; Putnam, F.W.

Proc. Natl. Acad. Sci. U.S.A. 81, 390-394, 1984

A:Title: Single-chain structure of human ceruloplasmin: the complete amino acid seq

A:Reference number: A00524; MUID:84119453; PMID:6582496

A:Accession: A00524

A:Molecule type: protein

A:Residues: 20-1060; 1065-1069 <YAN>

A:Note: 79-Gly and 449-Gly were also found

R:Yang, F.; Naylor, S.L.; Lam, J.B.; Cutshaw, S.; McCombs, J.L.; Naberhaus, R.H.;

Proc. Natl. Acad. Sci. U.S.A. 83, 3257-3261, 1986
 A:Title: Characterization, mapping, and expression of the human ceruloplasmin gene.
 A:Reference number: 159067; MID:86205876; PMID:3486416
 A:Accession: 159067
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 218-1069 <RES>
 A:Cross-references: GB:M1356; MID:q180248; PID:AAA51975.1; PID:q180249
 C:Comment: Ferroxidase is a blue, plasma alpha2-glycoprotein binding 6-7 copper ions per
 C:Comment: in Wilson's disease the plasma levels of ferroxidase are diminished or undetectable
 C:Comment: The three fragment chains are produced spontaneously during purification and
 C:Genetics:
 A:Gene: GDB:CP
 A:Cross-references: GDB:119069; OMIM:117700
 A:Map position: 3q23-q25
 A:Introns: 1006/3; 1061/1
 A:Note: the list of introns is incomplete
 C:Function:
 A:Description: catalyzes the oxidation of free iron(II) to iron(III) coupled with the re
 A:Note: Iron(II), but not Iron(III), is the form bound and transported by transferrin
 A:Note: other possible functions are amine oxidase activity, copper transport and homeos
 C:Superfamily: ferroxidase; ferroxidase repeat homology
 C:Keywords: acute phase; alternative splicing; copper; duplication; glycoprotein; oxidat
 F:1-19/Domain: signal sequence #status predicted <SIG>
 F:20-1069/Product: ferroxidase, long form #status experimental <MATL>
 F:20-1060/Product: ferroxidase, short form #status experimental <MATs>
 F:20-499/Product: ferroxidase repeat homology <FO1>
 F:23-357/Domain: ferroxidase repeat homology <FO2>
 F:373-718/Domain: ferroxidase repeat homology <FO3>
 F:501-905/Product: ferroxidase repeat homology <K50>
 F:733-1059/Domain: ferroxidase repeat homology <K50>
 F:907-1069/Product: ferroxidase repeat homology <K50>
 F:138-397/62/Binding site: carboxylate (Asn) (covalent) #status experimental
 F:174-200-276-357-534-560-637-718-874-900/Disulfide bonds: #status predicted
 F:1297-388-926/Binding site: carboxylate (Asn) (covalent) #status absent
 F:3297-338-343/Binding site: copper (His, Cys, His) (type 1) #status predicted
 F:358/Binding site: carboxylate (Asn) (covalent) (partial) #status experimental
 F:656-699-704-709/Binding site: copper (His, Cys, His, Met) (type 1) #status predicted
 F:994-1040-1045-1050/Binding site: copper (His, Cys, His, Met) (type 1) #status predicted

Query Match 10.6% Score 1320; DB 1; Length 1069;
 Best Local Similarity 19.2% Pred. No. 5-6e-61;
 Matches 398; Conservative 204; Mismatches 420; Indels 1054; Gaps 29;

```

OY      5 LSTCFPLCLRFCSATRRYLGAVLSMDYQSDLGE---LPLVDARPPRRPKSPFPPNT 61
DB      6 LGIFLFLCSTP-AWAKERKHYIGILFTWDY-ASDHGEKKLISVDHSHNITLQNGPURI 63
OY      62 SVYRKLTLEVEFTDHLFNIAKPRPPWGLGFTLOAEYDTVYITLKNASHPVSLAANG 121
DB      64 GRLYKKALYQYTDPEFRPTIEKRPVNLGPIGLIKAEKGDKVYVHLKNLASRPYTFHSHG 123
OY      122 VSYWASGAEVDQTSOREKEDDKVPGSGSHVYVQWLKNGPMASDPLCLTYSLSHV 181
DB      124 IYKKEHGAITPDNTIDFORADDKYRPEQOTYLLATEGSGSGDGNCTRYTHSHI 183
OY      182 DLVKDLSGLIGLIVCREGSLAKEKTYQL-HKFTLLPAVFDGSKWSHB-----TNS 234
DB      184 DAKDIASGLIGPLIKDLSKEKIDEFEVYVMSVVDENFMYLEDINIKTYCSEP 243
OY      235 LMODRDAASAPAMPKMTVNGYVNSLPGLIGCHRSYVHVHVGMTTEPVSHIFLEGHT 294
DB      244 EKVDKDNEDFQESNMRNVSVNGYTFGSLPGLSCADRYKWTLLGOMGNEVDVAHAFHQDA 303
OY      295 FLVRRNRQASLEISPTIFLTAQTLMDLQFLTFPHISSRHQHMNAVYKVSCEPEQL 354
DB      304 LTNKRNRTDTINLPATLFDAYVAONPGEWMLSQNLNHLKAGLQAFQVQBC----- 357
OY      355 RKNNEEADYDDDLTJSEMDVVRFPDDNSPSPFIQINSVAKKPKWVHVAIAEEDMDY 414
DB      358 ---NKSSKD-----NIRGKHVRH-----YIAAEETINNY 385
OY      415 APL-----VLAPDDBRSYKSOYLNNGPQIRIGKRYKRVFMAVYTDPEF---KTRERAIQ 462

```

```

DB      386 APSGDIPTFKENTLTAAGSDS--AVFEQGGTRRIGGYSKRYLYREYTDASLFNKRKGPPE 443
OY      463 HESGICPLLYGVEGVTLLIIRKNQASRPYNIYPHGI-----TDVRPLY---SRRLP 511
DB      444 EHLGIDLPYMAVGVGTLIVTFPHNKGAVPLSIEPIGVRRKNKNETIYSPYANQSRNPV 503
OY      512 KGVKHLDFPILGELFEKFKYTYVEGQFPKSPRCLLTRYSSFPNMEDLASGLIGPL 571
DB      504 PSASH-----VAFTETFEYEMVPEVGPNTADVCIAKMTYSADVPTKIDLETGLGPMK 558
OY      572 ICYKESVDQKQIMSDKRNVTLSFVEDENRSMYLTENTQFLPPNPAVOLDEFEQASN 631
DB      559 ICKKGSILHANGRQKDYKKEFYLPPEVDENSSLIDENIRMFETAPDQVDEDEDFQESN 618
OY      632 IMHSINGVYFDSIQ-LSVCLHEVAWYTLISGAQTDPLSVFPGYTFKKNVYEDTLTLE 690
DB      619 KHSNNGPFGNPGILNCKGDSVYVTLFSGNENADVHGILYFSGWYILMRGERDANLE 678
OY      691 PPSGETYFMSKNPGLMILGCHNSDFRRKGTALLKYSCDKNTQDYEDSYEDISALL 750
DB      679 PQTSLTHMMPDTEGFVNECLTDTHTYTGKKQKTYVNO----- 718
OY      751 SKNNALIEPRSPQNSRHPSTROKOPNATIPENDIEKTPWFAHRTPMKQIONVSSDLL 810
DB      719 ----- 718
OY      811 MLRSPPTPHGLSLDLEAKYETPSDDPSGALDSNNSLSEWTHRPQLHHSQDWVETP 870
DB      719 ---RRSE-----DSFYLYGERTY----- 735
OY      871 ESGLOLRNKLGTATATELKLIDRKVSTSNLITSPDNIAAGTNTSSLOPPMPV 930
DB      736 -----IAAVE----- 742
OY      931 HYDSQDITLFGKKSPLTESGGLSLEENNDKLLSGLNNSQSSMGKNVSTESGR 990
DB      743 ----- 742
OY      991 LFKGRAGPALTLKDNALFKVSLTKNTKSNSTNKTTHIDPSILLIENSPWQON 1050
DB      743 -----WD- 744
OY      1051 ILESPTEEKKATPLIHDRMLMDKNATALRLHNSKNTTSSKNMEVQKKSGPIPPDAON 1110
DB      745 ----- 744
OY      1111 PMSFFKMLLPESARMIQTHGKNSLNSGGSPKQVLSGPEKSVBGNPLSEKNKVY 1170
DB      745 ----- 744
OY      1171 VQGEFTKDVGLKEMVPPSSRNLLFLNLDNLHNNTHNOKKIOETIEKKTLLIQENVVL 1230
DB      745 ----- 744
OY      1231 PQIHTVYTKTFMKNLFLSTQNVGSDYDAVAPVLDPRSLNDSNTTKKHAHPSK 1290
DB      745 -----TSP----- 747
OY      1291 GEENLEGLNQTKQIYEKYACTRISPSNQOFTVQSRKALQFLPLETELEKRI 1350
DB      748 -----QR----- 749
OY      1351 IYDDPSQWKNMKHLPPSLTQIDVNEKKGALQSPSLDCLTRSHSIPQANSPLPIA 1410
DB      750 -----EMEKELNHL----- 758
OY      1411 KVSSEPSIRPIYTLRVLPQDSSHLPAASYRKKGSGVQESSHFLQAKANNLSIALITLE 1470
DB      759 -----QONVSNAPL----- 768
OY      1471 MTGDQREVSLGTSATNSVYKRVENVTLPRDLPKTSGLVEILLPVHITQKDLFPTETS 1530

```

Db 769 ----- 768
OY 1531 NGSPGHLDLVBSLLQTEGAIRKWNENARPGKPELRYATESAKTPESKILDPAMDNRH 1590
Db 769 ----- 768
OY 1591 GTOIPKEWKSQESPEKTAFFKKNITLISLACSNHAIALINEGKPEIETWAKOGR 1650
Db 769 ----- 768
OY 1651 TERLCSONPVLKRHQRREITFTLQSDQREIDYDTISVEKKKEDFDIYDEDNQSPSF 1710
Db 769 ----- 772
OY 1711 OKTRHFTIAYERLMDYGMSSPHYLARNAGSGVPOFKVYQOETDOSTFOPLRYGE 1770
Db 773 ----- 800
OY 1771 LNEHLGLGPYIRAEVEDNIWTERNOASPRYSFSSLSISTEEDOROGAEPKRNFEK 1827
Db 801 EEEHLGLGPOLHADYGRKALIFKNNATRPYSIHA-----HGVTESSTVTPRL 850
OY 1828 PNETKTYEKKVOHNAPTKDEFCANAYESDYLEKDVHSLIGPLVCHTNTLNPAG 1887
Db 851 PGFTLYVWKIPERSGAGTEDSACIPWAYYSFVDQVNDLYGLIGPLVCRRPYLKEVP 910
OY 1888 ROYVVOEFLAFTIDETKSWYFTEENNERKCAPQICQMEDPTKERYRPHAINGYIMDT 1947
Db 911 RRLK--EFALPLVYDENESMTLDNKTYSDBPEKYNKDEEFTESKHAINGMGN 968
OY 1948 LPLGLVMAQDRIKWTLSKSGNENISHPGSHVTKREKBYKALYNLYPOVFETVEM 2007
Db 969 LQGLTMVGDENVWYMLKNGNEIDLHGVHGHGSFYKRGVYSSDVFIDIFPGYOTLEM 1028
OY 2008 LPSKAGIWRVECLIGEHHLAQMSTLFLYVSNKQCP 2043
Db 1029 FPRTPGIMLHCHVTDIHAGMETTYVVLQNEGEYP 1064

RESULT 9
A:35210
N:Altername names: ceruloplasmin
C:Species: Rattus norvegicus (Norway rat)
C>Date: 10-Sep-1999 #sequence:Jewison10-Sep-1999 #text:change 10-Sep-1999
C:Accession: A35210; A41753; A29564; S21692
J: Biol. Chem. 265, 7701-7707, 1990
R: Fleming, R.E.; Gilfill, J.D.
A:Title: Primary structure of rat ceruloplasmin and analysis of tissue-specific gene expression
A:Reference number: A35210; PMID:90237081; PMID:2332446
A:Status: not compared with conceptual translation
A:Accession: A35210
A:Molecule type: mRNA
A:Residues: 1-1059 <FL>
A:Cross-references: GB:J05424
R: Fleming, R.E.; Gilfill, J.D.
J: Biol. Chem. 267, 479-486, 1992
A:Title: Structural and functional analysis of the 5'-flanking region of the rat ceruloplasmin gene
A:Reference number: A41753; PMID:92112697; PMID:1730611
A:Accession: A41753
A:Molecule type: DNA
A:Residues: 1-48 <FL2>
A:Note: the authors translated the codon GAA for residue 40 as Gly, GAA for residue 41 as Ala, and the authors translated the codon GCG for residue 60 as Gly and GGC for residue 1 as Alanine.
A:Experimental source: liver
A:Note: the authors translated the codon GCG for residue 60 as Gly and GGC for residue 1 as Alanine.
R: Ryan, T.P.; Grover, T.A.; Aust, S.D.

Arch. Biochem. Biophys. 293, 1-8, 1992
A:Title: Rat ceruloplasmin: resistance to proteolysis and kinetic comparison with human ceruloplasmin
A:Reference number: S21692; PMID:92117681; PMID:1531003
A:Accession: S21692
A:Molecule type: protein
A:Residues: 20-29, 'Q', 902-910 <RVA>
C:Superfamily: ferroxidase; ferroxidase repeat homology
C:Keywords: copper; glycoprotein; oxidoreductase; plasma
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-1059/Product: ferroxidase repeat homology <RVA>
F:23-356/Domain: ferroxidase repeat homology <RVA>
F:372-712/Domain: ferroxidase repeat homology <RVA>
F:727-1053/Domain: ferroxidase repeat homology <RVA>

Query Match 10.28; Score 1270; DB 1; Length 1059;
Best Local Similarity 19.08; Fred. No. 2.3e-58;
Matches 394; Conservative 196; Mismatches 425; Indels 1054; Gaps 29;

5 LSTCEFLCLRFCSAVRRYIGAVELSDYMQ-SDEL-PYDARFPPVKKSEPPNTS 62
6 LSLALFL-HSLAATREKHYYIGTEAWDASGEKELISVDEOSFYLRNGPDRIG 64
63 VYKRTLEFEETDLFNIAKRPWAGLGPITQAEYDVTYVITLNNASHPVSLHAGV 122
64 RYKRYALYSEYTDGTFKTKIDKPAWLGFLGPIYAEVGDVAVHKNFASRPYFHAHV 124
123 SVKASGEAEVDQTSOREKEDKVPKGSSTYVYVWYKENGPMASDPICLYLSHYD 182
125 TYKNAEGNIPDVTIDPQADKLPFGQYLYVRA-NPSPGEGDSNCTYRIYSHVD 183
183 LYKDLNSLIGALLVCREGSLAKETQTL-HKFTLLFAVDECKSHSEYKNSLM----- 226
184 APDIALSGIPLILCKKSLHKEKENDIDEFLMYSVDENLSEYEDNKTFCSEPE 243
237 -DPRDASARAMPKATVNGYVNRSLPGLIGCHRSYVWYHGMGTPEVSHLEGTE 295
244 KYDKMEDFOESNRMYSINGYTGSLPGISMCADBRKAVLFGNGEVYSHALFHCAL 303
296 LVNHRQASLEISPTPLTQTLMDLGOFLFCHISHODMEVAVVSCPEPQLR 355
304 TSKNYHTDILNFPATLIDVSNVQAPGVWMLSCGNLHAKGLAFVYRNC----- 356
356 MKNNEAEYDDLDSEMDVAFDDNNSPISQIRVAKKHKTAVYIAEEDMDYA 415
357 ---KNSP--DDDIQDNRV-----RH-----YINAEETIMDYA 385
416 P-----LVLPDDRSYSQYTLNNGPDRIGKRYKRVPMAYTDEF---KTEBAI 461
386 PGGTDFPGENLTSLGSDSVFPEQ---GATRIIGSGYKRLVREYTDSPFNKREKRPD 441
462 QHESGILGLYGEVDTLLIFKNOASPRYNIYPHGITDVR-----PLYSRLPFGYVHL 517
442 EHHIGLGPVIAEAGDIIRVTPHNKGPPSLISOPMGVRFKNEEGTYG---PDGRSK 498
518 KDFPILGPIFYKWTYVVEDGPKRDPKCLRRYSSPVNMHEDLNSLIGPLICYKES 577
499 QASHVAPKRTFTYEMVYKEMGPTVADPCLSKATYSVDLTIDITGLIGMKCKKGS 558
578 VDQRQOIMSDKRNVLIFSVPDENNSWYLTENIQRLFPNAGVQLEDPEFOASNMHSIN 637
559 LADGKQKNDKEFLATYVFDENESLILDDNIRFPTAENVDKDEDPQDSNMSHM 618
638 GYVDSIQ-LSYCLHFAVWYIISIAQDTFVSFSGSYFKRMVYEDTLTLPFSGET 656
619 GFYAGNLPCLNMGLOESIVYVLSAENADVIGIYSGNNTYLSKGRBDTANLFPKSLT 678
697 VPMSPNPMIWLIGCHNSDPNNGKPMALLKYSQCKNTGQYVDSYEDTISAVLSKNAI 756
679 LMTPTBESSEVVECLTIDHTTGAMQKTYVQC-----KGEEDVT----- 720
757 EPRFSQNRHPSRQKOFNATITIPENDIEKTDWPAHRTPMERIONVSSDMLMLROS 816
721 ----- 720

```

OY      817 PTPGLSLDLOQAKYETFSDDPFGAIDSNNSLSEKTHRPQLHHSQDMVTFESGLQL 876
DB      721 ----- 720
OY      877 RLNEKLGTTATELKLKIDRKVSTSNLITITPSDNLAAQDNTSILGPPSMFHYHSOL 936
DB      721 ----- 720
OY      937 DTTLEGGKSSPTLSEGGPLSLSENNDSKLEGLMNSQESSWGKNVSTESGRLFKGR 966
DB      721 ----- 726
OY      997 AHGPALTLKQMALEKVISILKTNKTSNNSATNKRTHIDGPIILBNPSVWONLESDF 1056
DB      727 TYYTIA----- 732
OY      1057 EFKKVTPLIDRMLMDKNATALLAHNSKNTTSKKNEMVQCKKSGPIPPDQNPDMSPFF 1116
DB      733 ----- 732
OY      1117 KMLFLPESARWIOPTHGKNSLNSGGSPSPKQVSLGPEKSEVQNFLEKKNVYVKGGEF 1176
DB      733 ----- 732
OY      1177 TKDVLKEMVFPSSRNLETLNLDLHNNHTHNOEKKIQEIEKEKTELIOENVLPOLHTV 1236
DB      733 ----- 732
OY      1237 TGTAKFMKNLELSTRQNVGSGIDGAYAPVQDPRSLNDSTNTRKHTAHPSKGEENL 1296
DB      733 ----- 741
OY      1297 EGIHQTKQIVKACTRTISPTSQNFQORSKRALKQFLPLETELEKRIIVDDTS 1356
DB      742 -----S 742
OY      1357 TOMSKNMKHLTPSTLQIDYNEKEKALITOSPDISCITRSHSTIPQANRSPILIAKVSSFP 1416
DB      743 RDMEMELHHL----- 752
OY      1417 STRPIYLRVLFQDNSSHLPAASYRKKDQSGVOESHSFLQAKKNNLSLILTEMGQDR 1476
DB      753 -----QDNVSNAPL----- 762
OY      1477 EVGSLGTSATNSVYTKKVENTVLPKPDLPKTSQKVELLPVHIYQKDLFPETESNGSPGH 1536
DB      763 ----- 762
OY      1537 LDVGGSLQGTGALIKMNEANRPGKVPPLRVATESAKTPSKLIDPLAMDNHYGTQIPK 1596
DB      763 ----- 762
OY      1597 EEMKSGEKSPEKTAFFKKKOTIILSNACESNHAIAINEGONKEPELEVWAKGRTERLCS 1656
DB      763 ----- 762
OY      1657 QNPPVILKRQRETRTTLQSDQEEIDYDTLISEKKEDFDIYDEBNSPSPFOKTRH 1716
DB      763 -----DKEE----- 766
OY      1717 YETIAVERLMDYGMSSPHVILRRNRAQSGSVPOFKVVEQEFIDQSFLOPLYGELNEHLG 1776
DB      767 FFIQS-----KYKVVYREFIDSTFREQVKRRAREEHIG 800
OY      1777 LIGPYIRAEVDNIMTFERNQASRPYSLSLISEEDROGAPPRKNVYK---PNEIKT 1833
DB      801 ILDEPLIHADYQAKVAVYFKMATRPYSIHA-----HGVTKSYIYAPLLEGEVKT 850
OY      1834 YFWKVOHNAPTKDEFDCKAMAYFSVDLEKDVHSLIGLPLVC---HTNLTNPAHSGOV 1890
DB      851 YIMQIPERSGAGTEDSPICPMAYSTVDKYLIGLIGLIPLYCKRSKAVYKPNK----- 905

```

```

OY      1891 TVQBFALFTIEDKTSWYFTEENNERNCRAPNIQNEDETFEKENTRPHAINGYIKOTLPG 1950
DB      906 KMEFSLILEYEDENSWTLDNINTYDPDEKVENNDNEEFIESKMAIANGKFGNLQ 965
OY      1951 LVMAODRIRWYLLSGNSNENHSIHSFGHFTVRKEEYKAKALNYLPGVFETVELPS 2010
DB      966 LHMVGDVWMMYMAAGNIDILHVFHHRHGSFQKIRGLSHSDVDPEFGYOTLEMPQ 1025
OY      2011 KAGTWRVCELTGEHLHAGMSTFLYVSNK 2039
DB      1036 TPGTWLHCHVTDHIRAGVTTYVLPNQ 1054

```

```

RESULT 10
A44258
factor VIIr-associated gene B hypothetical protein - human
C:Species: Homo sapiens (man)
C>Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 18-Jun-1999
C:Accession: A44258
C:Levinson, B.; Kenwright, S.; Gamel, P.; Fisher, K.; Gitschler, J.
Genomics 14, 385-389, 1992
A>Title: Evidence for a third transcript from the human factor VIII gene.
A:Reference number: A44258; M0ID:93052386; PMID:1427867
A:Accession: A44258
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-216 <LEV>
A:Cross-references: GB:M90707; NID:g182316; PIDN:AA58466.1; PID:g182317
C:Superfamily: coagulation factor VIIIr, discoidin I amino-terminal homology; ferrox;
F:1-53/Domains: discoidin I amino-terminal homology #status atypical <DN1>
F:57-210/Domains: discoidin I amino-terminal homology <DN2>

```

```

Query Match      8.9%: Score 1104; DB 2; Length 216;
Best Local Similarity 100.0%: Pred. No. 8.3e-51;
Matches 208; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      2144 VFGNVDSSGIGKHINFPPIARVIRLPHYTSIRSTLWELMGDLNSGMPGMEKSA 2203
DB      9 VFGNVDSSGIGKHINFPPIARVIRLPHYTSIRSTLWELMGDLNSGMPGMEKSA 68
OY      2204 ISDAQITASSYFTNMFATSPSKARLHLQGSNAMPQVNNPKEMLOYDFQTKKVTGYT 2263
DB      69 ISDAQITASSYFTNMFATSPSKARLHLQGSNAMPQVNNPKEMLOYDFQTKKVTGYT 128
OY      2264 TQGVKSLTSMTYKEFLISSQDGHQWLFQNKYKQVQNDSPFPVNSLDPPLTR 2323
DB      129 TQGVKSLTSMTYKEFLISSQDGHQWLFQNKYKQVQNDSPFPVNSLDPPLTR 188
OY      2324 YLRHPOSVHQIALRMEYLGCEADLY 2351
DB      189 YLRHPOSVHQIALRMEYLGCEADLY 216

```

```

RESULT 11
JC4915
ags protein precursor - rat
N:Alternate names: O-acetyl-Gd3 ganglioside.
C:Species: Rattus norvegicus (Norway rat)
C>Date: 26-Sep-1996 #sequence_revision 01-Nov-1996 #text_change 20-Jun-2000
C:Accession: JC4915
C:Ogura, K.; Nara, K.; Watanabe, Y.; Kohno, K.; Tai, T.; Sanae, Y.
Biochem. Biophys. Res. Commun. 225, 932-938, 1996
A>Title: Cloning and expression of cDNA for O-acetylation of GD3 ganglioside.
A:Reference number: JC4915; M0ID:96374422; PMID:8780713
A:Accession: JC4915
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-427 <OGU>
A:Cross-references: DDBJ:D84068; NID:g1620006; PIDN:BA12210.1; PID:g1620007
A:Experimental source: GST cell
C:Comment: This protein is required for the O-acetylation of disialoganglioside sI
A:Gene: ags

```


A>Title: Molecular cloning of glycoprotein antisera MGP57/53 recognized by monoclonal an
A:Reference number: S65138; MUID:96125736; PMID:8541316
A:Accession: S65138
A>Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-401 <AOK>
A:Keywords: I.H.; Banghart, L.R.; Lane, W.S.
Biochem. Mol. Biol. Int. 29, 545-554, 1993
A>Title: The major fat globule membrane proteins, bovine components 15/16 and guinea pig
IT-like sequences
A:Reference number: A48394; MUID:93250576; PMID:8465470
A:Accession: G48394
A>Status: preliminary
A:Molecule type: protein
A:Residues: 207-220 <MAT>
A:Experimental source: milk
A>Note: sequence extracted from NCBI backbone (NCBI:P131457)
C:Superfamily: milk fat globule protein; discoidin I amino-terminal homology; EGF homolo
C:Keywords: glycoprotein
F:1-32/Domain: EGF homology <EG2>
F:40-79/Domain: EGF homology <EG2>
F:83-239/Domain: discoidin I amino-terminal homology <DN2>
F:243-401/Domain: discoidin I amino-terminal homology <DN2>

Query Match 5.1%; Score 635; DB 2; Length 401;
Best Local Similarity 37.9%; Pred. No. 8e-26;
Matches 135; Conservative 66; Mismatches 123; Indels 32; Gaps 7;

QY 2018 ECLIGELHAGMSTFLVYSNK-----COTPLGASGHINDPQTASGQY-- 2062
DB 50 ECQVYDSDHSG--DVFQYICCKPLGYVGHCECTCTSPGMDGALDSQISASMHG 107
QY 2063 -----GMAPKIARLHYSGSINAMST-----KEPESWKVDLLAPMTHIGIKTGAROKFSS 2114
DB 108 FNGLOKMAPELARLHQTGIYNAWTSNDKDP--WIOVNMKRMWVTGVYVQASRAGSA 165
QY 2115 LYSQFIIMSLDGKMKQYRNGSTGLMVFPGVNDSSGKHNIFNPPIARYRLPETH 2174
DB 166 EYLKTFKVAYSTGRQOFIYVAGSGDKIFIGVNNSSGKINLDFPLETYVLPYII 225
QY 2175 YSIRSTLMELMGCDLNSCMPLGESAKISDAQITASSYFTN--MPTWSPSKARLHL 2231
DB 226 CHRGTLEFELGCELGCTEPLGLKNDWTIPNKQITASSYTKTWGLSAPFPYARLDN 265
QY 2232 QGRSNAMPQVNNKPEMLQVDPQKTMKYGTGTVTGVSLLTSMYKEFLISSODGQMT 2291
DB 286 QGRFNAMTAQTNSASEMLQIDLSQKRYTGITGADDFGHIOYAAVAVAGDGVYWT 345
QY 2292 LFFONG--KRVVFGNDSFTPVNSLDPPLLTRYLRIHQSVWHDALRMEVLGC 2345
DB 346 EYKDPGASESKIFPGNMDNNSHKNIFFETPQAFVRIQVAMNHRITLRVELLGC 401

RESULT 15

S74211

PAS-6/7 protein precursor - bovine

C:Species: Bos primigenius taurus (cattle)

C:Date: 04-Dec-1997 #sequence revision 12-Dec-1997 #text change 20-Jun-2000

C:Accession: S74211; S78114; S24181

R:Hvartregaard, J.; Andersen, M.H.; Berglund, L.; Rasmussen, J.T.; Petersen, T.E.

Eur. J. Biochem. 240, 628-636, 1996

A>Title: Characterization of glycoprotein PAS-6/7 from membranes of bovine milk fat glob

A:Reference number: S74211; MUID:97008954; PMID:8856064

A:Accession: S74211

A:Molecule type: mRNA

A:Residues: 1-427 <HVA>

A:Cross-references: EMBL:X91895; NID:91632778; PIDD:CAA62997.1; PTD:91632778

A:Accession: S78114

A:Molecule type: protein

A:Residues: 19-85;96-110;140-165;174-216;221-232;248-277;285-293;309-337;339-420;425-427

R:Klein, D.H.; Kanno, C.; Mizokami, Y.

Biochim. Biophys. Acta 1122, 203-211, 1992

A>Title: Purification and characterization of major glycoproteins, PAS-6 and PAS-7, from

A:Reference number: S23926; MUID:92353107; PMID:1643094
A:Accession: S24181
A:Molecule type: protein
A:Residues: 383-394 <KIN>
C:Superfamily: milk fat globule protein; discoidin I amino-terminal homology; EGF ho
C:Keywords: blocked amino end; disulfide bond; glycoprotein; milk
F:1-18/Domain: signal sequence #status predicted <SIG>
F:19-427/Product: PAS-6/7 protein #status experimental <MAT>
F:124-58/Domain: EGF homology <EG2>
F:108-265/Domain: EGF homology <EG2>
F:269-427/Domain: discoidin I amino-terminal homology <DN2>
F:24-35;29-47;49-56;66-77;71-94;96-105/Disulfide bonds: #status predicted
F:27/Binding site: carbohydrate (Ser) (covalent) #status experimental
F:34/Binding site: carbohydrate (Thr) (covalent) #status experimental
F:59;227/Binding site: carbohydrate (Asn) (covalent) #status experimental
F:109-265;252-256;270-427/Disulfide bonds: #status experimental

Query Match 5.1%; Score 635; DB 2; Length 427;
Best Local Similarity 37.9%; Pred. No. 8.8e-26;
Matches 135; Conservative 66; Mismatches 123; Indels 32; Gaps 7;

QY 2018 ECLIGELHAGMSTFLVYSNK-----COTPLGASGHINDPQTASGQY-- 2062
DB 76 ECQVYDSDHSG--DVFQYICCKPLGYVGHCECTCTSPGMDGALDSQISASMHG 133
QY 2063 -----GMAPKIARLHYSGSINAMST-----KEPESWKVDLLAPMTHIGIKTGAROKFSS 2114
DB 134 FNGLOKMAPELARLHQTGIYNAWTSNDKDP--WIOVNMKRMWVTGVYVQASRAGSA 191
QY 2115 LYSQFIIMSLDGKMKQYRNGSTGLMVFPGVNDSSGKHNIFNPPIARYRLPETH 2174
DB 192 EYLKTFKVAYSTGRQOFIYVAGSGDKIFIGVNNSSGKINLDFPLETYVLPYII 251
QY 2175 YSIRSTLMELMGCDLNSCMPLGESAKISDAQITASSYFTN--MPTWSPSKARLHL 2231
DB 252 CHRGTLEFELGCELGCTEPLGLKNDWTIPNKQITASSYTKTWGLSAPFPYARLDN 311
QY 2232 QGRSNAMPQVNNKPEMLQVDPQKTMKYGTGTVTGVSLLTSMYKEFLISSODGQMT 2291
DB 312 QGRFNAMTAQTNSASEMLQIDLSQKRYTGITGADDFGHIOYAAVAVAGDGVYWT 371
QY 2292 LFFONG--KRVVFGNDSFTPVNSLDPPLLTRYLRIHQSVWHDALRMEVLGC 2345
DB 372 EYKDPGASESKIFPGNMDNNSHKNIFFETPQAFVRIQVAMNHRITLRVELLGC 427

Search completed: July 2, 2003, 09:47:00
Job time: 90 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2003, 09:32:20 ; Search time 33 Seconds

(without alignments)
2954.874 Million cell updates/sec

Title: NP_000123

Perfect score: 12418
Sequence: 1 mqlstcfcflcllrfaa.....vwhqalrmevjgcaagdy 2351

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Swisprot_40:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Query Length	DB ID	Description
1	12418	100.0	2351	1	FA8_HUMAN
2	8835	71.1	2319	1	FA8_MOUSE
3	8076	65.0	2133	1	FA8_PIG
4	2816.5	22.7	2224	1	FA5_HUMAN
5	2774.5	22.3	2258	1	FA5_PIG
6	2744	22.1	2211	1	FA5_BOVIN
7	1316	10.6	1065	1	CERU_HUMAN
8	1267	10.2	1059	1	CERU_RAT
9	1236	10.0	1062	1	CERU_MOUSE
10	663	5.3	427	1	MEGM_RAT
11	657	5.3	463	1	MEGM_MOUSE
12	650	5.2	409	1	MEGM_PIG
13	635	5.1	427	1	MEGM_BOVIN
14	588	4.7	387	1	MEGM_HUMAN
15	469.5	3.8	931	1	MRP2_HUMAN
16	464.5	3.7	925	1	MRP2_RAT
17	462.5	3.7	931	1	MRP2_MOUSE
18	458.5	3.7	914	1	MRP1_CHICK
19	451.5	3.6	922	1	MRP1_RAT
20	446.5	3.6	923	1	MRP1_MOUSE
21	443	3.6	928	1	MRP1_XENTIA
22	429.5	3.5	923	1	MRP1_HUMAN
23	306.5	2.5	3133	1	HMCT_BOMBO
24	260.5	2.1	280	1	MRP1_FGFR
25	247	2.0	224	1	MRP1_MOUSE
26	243	2.0	224	1	MRP1_HUMAN
27	242.5	2.0	3418	1	MRP2_HUMAN
28	235.5	1.9	1358	1	MRP2_MOUSE
29	218.5	1.8	1928	1	MRP2_YEAST
30	218.5	1.8	3329	1	MRP2_MOUSE
31	217	1.7	1331	1	CTA2_HUMAN
32	217	1.7	1381	1	YRPF_YEAST
33	216.5	1.7	1271	1	Y338_MYC

34	212	1.7	1957	1	YD86_SCHPO
35	212	1.7	3924	1	ANK2_HUMAN
36	211	1.7	1284	1	NRX4_DROME
37	209.5	1.7	1177	1	Y307_MYC
38	209	1.7	1420	1	APX_XENTIA
39	208	1.7	1310	1	CTA4_MOUSE
40	206	1.7	1308	1	CTA4_HUMAN
41	204	1.6	1165	1	YRPF_YEAST
42	203.5	1.6	1189	1	YRPF_YEAST
43	203.5	1.6	1762	1	DPOC_HUMAN
44	203	1.6	2245	1	MRP1_DICDI
45	202	1.6	6669	1	NEBU_HUMAN

ALIGNMENTS

RESULT 1	FA8_HUMAN	STANDARD:	PRT: 2351 AA.
AC	P00451:		
DT	21-JUL-1986 (Rel. 01, Created)		
DT	21-JUL-1986 (Rel. 01, Last sequence update)		
DT	15-JUN-2002 (Rel. 41, Last annotation update)		
DE	Coagulation factor VIII precursor (Procoagulant component)		
DE	(Antihemophilic factor) (AHF).		
GN	F8 OR F8C.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=6081154; PubMed=3935400;		
RA	Tinetti M.A., Blacher R., Burke R.L., Caput D., Chu C., Dina D.,		
RA	Harlog K., Kuo C.H., Mastarz F.R., Merryweather J.P., Majzarian R.,		
RA	Pachl C., Potter S.J., Puma J., Quiroga M., Rall L.B., Randolph A.,		
RA	Urdea M.S., Valenzuela P., Dahl H.-H.M., Favalario J., Hansen J.,		
RA	Nordfang O., Eban M.;		
RT	*Characterization of the polypeptide composition of human factor		
RT	VIII: C and the nucleotide sequence and expression of the human kidney		
RT	cDNA. *		
RL	DNA 4:333-349(1985).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=85061548; PubMed=6438526;		
RA	Wood W.I., Capon D.H., Simonsen C.C., Eaton D.L., Gitschler J.,		
RA	Keyt B., Seeburg P.H., Smith D.H., Hollingshead P., Wion K.L.,		
RA	Delwart B., Tuddenham E.G.D., Vehar G.A., Lawn R.M.;		
RT	*Expression of active human factor VIII from recombinant DNA clones. *		
RT	Nature 312:330-337(1984).		
RN	[3]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=85061550; PubMed=6438528;		
RA	Toole J., Knopf J.L., Wozney J.M., Sultzman L.A., Buecker J.L.,		
RA	Pittman D.D., Kaufman R.J., Brown E., Shoemaker C., Orr E.C.,		
RA	Amphlett G.W., Foster W.B., Coe M.L., Knutson G.J., Fass D.N.,		
RA	Hewick R.M.;		
RT	*Molecular cloning of a cDNA encoding human antihemophilic factor. *		
RT	Nature 312:342-347(1984).		
RN	[4]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=93265012; PubMed=1303178;		
RA	Gitschler J., Wood W.I.;		
RT	*Sequence of the exon-containing regions of the human factor VIII		
RT	gene. *		
RL	Hum. Mol. Genet. 1:199-200(1992).		
RN	[5]		
RP	SEQUENCE OF 2064-2070 FROM N.A.		
RA	de Water N.S., Williams R., Brownett P.J.;		
RL	Submitted (JUN-1997) to the EMBL/GenBank/DBD databases.		
RN	[6]		
RP	SOLUTION OF TYR-1699.		

RX MEDLINE-91093266; PubMed-1898735;
 RA Leyte A., van Schijndel H.B., Niehrs C., Huttner W.B., Verbeet M.P.,
 RA Mertens K., van Mourik J.A.;
 RT "Sulfation of Tyr1680 of human blood coagulation factor VIII is
 RT essential for the interaction of factor VIII with von Willebrand
 RT factor.";
 RL J. Biol. Chem. 266:740-746(1991).
 RN [17]
 RP SULFATION.
 RX MEDLINE-92207952; PubMed-1554716;
 RA Pittman D.D., Wang J.H., Kaufman R.J.;
 RT "Identification and functional importance of tyrosine sulfate
 RT residues within recombinant factor VIII.";
 RL Biochemistry 31:3315-3325(1992).
 RN [18]
 RP STRUCTURE BY NMR OF 2322-2343.
 RX MEDLINE-95200924; PubMed-7893714;
 RA Gilbert G.E., Baleja J.D.;
 RT "Membrane-binding peptide from the C2 domain of factor VIII forms an
 RT amphipathic structure as determined by NMR spectroscopy.";
 RL Biochemistry 34:3022-3031(1995).
 RN [19]
 RP REVIEW ON MOLECULAR BASIS OF HEMOPHILIA A.
 RX MEDLINE-91221499; PubMed-1902642;
 RA Gitschler J.;
 RT "The molecular basis of hemophilia A.";
 RL Ann. N.Y. Acad. Sci. 614:89-96(1991).
 RN [10]
 RP REVIEW ON MOLECULAR BASIS OF HEMOPHILIA A.
 RX MEDLINE-8908506; PubMed-2491949;
 RA White G.C. II, Shoemaker C.B.;
 RT "Factor VIII gene and hemophilia A.";
 RL Blood 73:1-12(1989).
 RN [11]
 RP REVIEW ON MOLECULAR BASIS OF HEMOPHILIA A.
 RX MEDLINE-95245332; PubMed-7728145;
 RA Antonarakis S.E., Kazazian H.H., Tuddenham E.G.D.;
 RT "Molecular etiology of factor VIII deficiency in hemophilia A.";
 RL Hum. Mutat. 5:1-22(1995).
 RN [12]
 RP VARIANT GLN-2326.
 RX MEDLINE-86235434; PubMed-3012775;
 RA Gitschler J., Wood W.I., Shuman M.A., Lawn R.M.;
 RT "Identification of a missense mutation in the factor VIII gene of a
 RT mild hemophilic.";
 RL Science 232:1415-1416(1986).
 RN [13]
 RP VARIANT PRO-2135.
 RX MEDLINE-86096539; PubMed-3122181;
 RA Levinson B., Janco R.L., Phillips J.A. III, Gitschler J.;
 RT "A novel missense mutation in the factor VIII gene identified by
 RT analysis of amplified hemophilia DNA sequences.";
 RL Nucleic Acids Res. 15:9797-9805(1987).
 RN [14]
 RP VARIANT GLN-2228.
 RX MEDLINE-88191889; PubMed-2833855;
 RA Youssoufian H., Antonarakis S.E., Bell W., Griffin A.M.,
 RA Kazazian H.H.;
 RT "Nonsense and missense mutations in hemophilia A: estimate of the
 RT relative mutation rate at CG dinucleotides.";
 RL Am. J. Hum. Genet. 42:718-725(1988).
 RN [15]
 RP VARIANT GLY-291.
 RX MEDLINE-88220354; PubMed-2835904;
 RA Youssoufian H., Wong C., Aronis S., Placoukis H., Kazazian H.H. Jr.,
 RA Antonarakis S.E.;
 RT "Moderately severe hemophilia A resulting from Glu-->Gly substitution
 RT in exon 7 of the factor VIII gene.";
 RL Am. J. Hum. Genet. 42:867-871(1988).
 RN [16]
 RP VARIANT CYS-1708.
 RX MEDLINE-89274393; PubMed-2499363;
 RA O'Brien D.P., Tuddenham E.G.;
 RT "Purification and characterization of factor VIII 1,689-Cys: a
 RT nonfunctional cofactor occurring in a patient with severe hemophilia
 RT A.";
 RL Blood 73:2117-2122(1989).
 RN [17]
 RP VARIANT CYS-391.
 RX MEDLINE-90001543; PubMed-2506948;
 RA Shima M., Ware J., Yoshikawa A., Fukui H., Fulcher C.A.;
 RT "An arginine to cysteine amino acid substitution at a critical
 RT thrombin cleavage site in a dysfunctional factor VIII molecule.";
 RL Blood 74:1612-1617(1989).
 RN [18]
 RP VARIANT LEU-189.
 RX MEDLINE-90057680; PubMed-2510835;
 RA Chan V., Chan T.K., Tong T.M., Todd D.;
 RT "A novel missense mutation in exon 4 of the factor VIII:C gene
 RT resulting in moderately severe hemophilia A.";
 RL Blood 74:2688-2691(1989).
 RN [19]
 RP VARIANT LEU-2326.
 RX MEDLINE-89197216; PubMed-2495245;
 RA Ioabba H., Fujimaki M., Kazazian H.H. Jr., Antonarakis S.E.;
 RT "Wild hemophilia A resulting from Arg-to-Leu substitution in exon 26
 RT of the factor VIII gene.";
 RL Hum. Genet. 81:335-338(1989).
 RN [20]
 RP VARIANT HIS-391.
 RX MEDLINE-89264602; PubMed-2498882;
 RA Arai M., Ioabba H., Higuchi M., Antonarakis S.E., Kazazian H.H. Jr.,
 RA Fujimaki M., Hoyer L.W.;
 RT "Direct characterization of factor VIII in plasma: detection of a
 RT mutation altering a thrombin cleavage site
 RT (arginine-372-->histidine).";
 RL Proc. Natl. Acad. Sci. U.S.A. 86:4277-4281(1989).
 RN [21]
 RP VARIANT CYS-1708.
 RX MEDLINE-90105723; PubMed-2104766;
 RA Arai M., Higuchi M., Antonarakis S.E., Kazazian H.H. Jr.,
 RA Phillips J.A. III, Janco R.L., Hoyer L.W.;
 RT "Characterization of a thrombin cleavage site mutation (Arg 1689 to
 RT Cys) in the factor VIII gene of two unrelated patients with
 RT cross-reacting material-positive hemophilia A.";
 RL Blood 75:384-389(1990).
 RN [22]
 RP VARIANT GLN-2228 AND LEU-2326.
 RX MEDLINE-90123183; PubMed-2105106;
 RA Casula L., Murru S., Pecorara M., Ristaldi M.S., Restagno G.,
 RA Mancuso G., Morfini M., de Biasi R., Baudo F., Carbonara A.;
 RT "Recurrent mutations and three novel rearrangements in the factor
 RT VIII gene of hemophilia A patients of Italian descent.";
 RL Blood 75:662-670(1990).
 RN [23]
 RP VARIANT CYS-391.
 RX MEDLINE-90329422; PubMed-1973901;
 RA Patlinson J.K., McVey J.H., Boon M., Ajani A., Tuddenham E.G.;
 RT "CRM+ hemophilia A due to a missense mutation (372-->Cys) at the
 RT internal heavy chain thrombin cleavage site.";
 RL Br. J. Haematol. 75:73-77(1990).
 RN [24]
 RP VARIANTS PRO-1699 AND CYS-1708.
 RX MEDLINE-90152651; PubMed-210506;
 RA Higuchi M., Wong C., Kochan L., Olek K., Aronis S., Kasper C.K.,
 RA Kazazian H.H., Antonarakis S.E.;
 RT "Characterization of mutations in the factor VIII gene by direct
 RT sequencing of amplified genomic DNA.";
 RL Genomics 6:65-71(1990).
 RN [25]
 RP VARIANTS CYS-1728 AND ASP-1941.
 RX MEDLINE-90169968; PubMed-2106480;
 RA Traystman M.D., Higuchi M., Kasper C.K., Antonarakis S.E.,
 RA Kazazian H.H.;
 RT "Use of denaturing gradient gel electrophoresis to detect point
 RT mutations in the factor VIII gene."

Query Match 100.0%; Score 12418; DB 1; Length 2351;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2351; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MQEISTCFCLLRFCFSATRRYYLGAVALSMDYMSDGLDELVDARPPRPKSPFPN 60
 1 MQEISTCFCLLRFCFSATRRYYLGAVALSMDYMSDGLDELVDARPPRPKSPFPN 60
 61 TSYYKKTLLPEFTDHLNINAKPRPPMGLGPTIOAEVYQTVITLKMAASPIYLAHV 120
 61 TSYYKKTLLPEFTDHLNINAKPRPPMGLGPTIOAEVYQTVITLKMAASPIYLAHV 120
 121 GYSTWAKASBGAELYDQTSQREKEDKVPGGSHYYVOYLKENGPMASDPLCLTYSYLSH 180
 121 GYSTWAKASBGAELYDQTSQREKEDKVPGGSHYYVOYLKENGPMASDPLCLTYSYLSH 180
 181 VDLVKDLSGLIGALLVCREGLAKERTQTLAKFTLLPAVDEGKSWHSFTKNSLMQDND 240
 181 VDLVKDLSGLIGALLVCREGLAKERTQTLAKFTLLPAVDEGKSWHSFTKNSLMQDND 240
 241 AASARAPKMHVYNGVYVRSLLPGLIGCRKSYVMHVIQGTTPYVHSIFLEGHTFLYRNH 300
 241 AASARAPKMHVYNGVYVRSLLPGLIGCRKSYVMHVIQGTTPYVHSIFLEGHTFLYRNH 300
 301 ROASLEISPTTFLTQOTLMDLGOELLFCHISHOHOGMEAVYVDSCPREPOLBKKNNE 360
 301 ROASLEISPTTFLTQOTLMDLGOELLFCHISHOHOGMEAVYVDSCPREPOLBKKNNE 360
 361 EAEDYDDDLTDSHMDVYRFDNDSPSTIQIRYAKKHRTVHYIYAAEEDMDYAPLVIA 420
 361 EAEDYDDDLTDSHMDVYRFDNDSPSTIQIRYAKKHRTVHYIYAAEEDMDYAPLVIA 420
 421 PDDRSYKSOYLNNGPORIGRKYKKRVMATYDTFETREATIOHESGILGLLYGEVDTL 480
 421 PDDRSYKSOYLNNGPORIGRKYKKRVMATYDTFETREATIOHESGILGLLYGEVDTL 480
 481 LIIFKNOASRPYNIYHGTIDVPRYLRPKVKHLKDFLLPGEIIFKKMTVYEDDP 540
 481 LIIFKNOASRPYNIYHGTIDVPRYLRPKVKHLKDFLLPGEIIFKKMTVYEDDP 540
 541 TKSDPRLTRVYSSPVMNERDLAGLIGPLITCYKESVDQKGNQIMSDKNVILLESVPE 600
 541 TKSDPRLTRVYSSPVMNERDLAGLIGPLITCYKESVDQKGNQIMSDKNVILLESVPE 600
 601 NRSWYLTENIQRLPYPAGVQLEDEPQASNIHASTNGTVFDSQLSYCLAEVAHYIILS 660
 601 NRSWYLTENIQRLPYPAGVQLEDEPQASNIHASTNGTVFDSQLSYCLAEVAHYIILS 660
 661 IGAQTDPLSVFSGYTFEKHKVYEDTLTLFPFSGETVFMSPENPLWTLICHSNDFRNG 720
 661 IGAQTDPLSVFSGYTFEKHKVYEDTLTLFPFSGETVFMSPENPLWTLICHSNDFRNG 720
 721 MTALLKVSQCDKNTGYEDSYEDISAVLLSKNNALEPVSFQNSRHPSRQKOPNATYI 780
 721 MTALLKVSQCDKNTGYEDSYEDISAVLLSKNNALEPVSFQNSRHPSRQKOPNATYI 780
 781 PENDIEKTDWFAHRPMPKIONVSSDGLMLROSPTPHGISLSDLOEKYETFEDDDS 840
 781 PENDIEKTDWFAHRPMPKIONVSSDGLMLROSPTPHGISLSDLOEKYETFEDDDS 840
 841 PGALDSNNLSSEWTHFRPOLHSGDMVTPPSGLQLRLNEKLGTAAATELKLDPKVSST 900
 841 PGALDSNNLSSEWTHFRPOLHSGDMVTPPSGLQLRLNEKLGTAAATELKLDPKVSST 900
 901 SNNLISITPSDNLAAQTDNTSSIGPPSPVHYVSQDITTLFCKSSPLTESGGSLISSE 960
 901 SNNLISITPSDNLAAQTDNTSSIGPPSPVHYVSQDITTLFCKSSPLTESGGSLISSE 960
 961 NNDSKLLESGLMNSQSSGKGNVSTESGRLFGRRAHGPALLTQDNALFVYSISLTKN 1020
 961 NNDSKLLESGLMNSQSSGKGNVSTESGRLFGRRAHGPALLTQDNALFVYSISLTKN 1020

1021 KTSNNSATNRKTHIDBSLLIENSFVQNTLESDFEKKVTPLLIHDBMLDKXNAPALRY 1080
 1021 KTSNNSATNRKTHIDBSLLIENSFVQNTLESDFEKKVTPLLIHDBMLDKXNAPALRY 1080
 1081 NMSKRTTSSKNEMVQOKKEGPIPPDAQNDQMSFFKMLFLESARAWIORTHGKNSLNSG 1140
 1081 NMSKRTTSSKNEMVQOKKEGPIPPDAQNDQMSFFKMLFLESARAWIORTHGKNSLNSG 1140
 1141 OGSPKOLVSLGPEKSYEGONFLSEKKNVYVVGKEEFGKQGLKEMVPPSSNRLFTYMLDN 1200
 1141 OGSPKOLVSLGPEKSYEGONFLSEKKNVYVVGKEEFGKQGLKEMVPPSSNRLFTYMLDN 1200
 1201 LBNNTNHOEKKIOELTEKKEITLQBNVYLQIHTYVGTAKMKULPLFSROVYESVD 1260
 1201 LBNNTNHOEKKIOELTEKKEITLQBNVYLQIHTYVGTAKMKULPLFSROVYESVD 1260
 1261 GAVAPVLDPRSLNDSTNRKTHAHFSKKEEENLGLQNOTKOIVERVACTTRISPNP 1320
 1261 GAVAPVLDPRSLNDSTNRKTHAHFSKKEEENLGLQNOTKOIVERVACTTRISPNP 1320
 1321 SOONFVTOSSKRALKOPRLPLETELEKRIIVDDTSTQMSKNNKHLTSTYLQIDYNEKE 1380
 1321 SOONFVTOSSKRALKOPRLPLETELEKRIIVDDTSTQMSKNNKHLTSTYLQIDYNEKE 1380
 1381 KGAITOSPCLDRSHSIPQANRSPPLAVYSSPISRIYTLRVLFDQDSSHLPAAS 1440
 1381 KGAITOSPCLDRSHSIPQANRSPPLAVYSSPISRIYTLRVLFDQDSSHLPAAS 1440
 1441 RKDGSVOESSHFLQGAKKNNLSLALILEMTGQREYVSLGTSATNSYTYKRYENTVLP 1500
 1441 RKDGSVOESSHFLQGAKKNNLSLALILEMTGQREYVSLGTSATNSYTYKRYENTVLP 1500
 1501 KPDLPKTSQKELLPRKVIYQKDLFPTETSNQSPGHLDLVEGSLQGTGEGAIKKNENARP 1560
 1501 KPDLPKTSQKELLPRKVIYQKDLFPTETSNQSPGHLDLVEGSLQGTGEGAIKKNENARP 1560
 1561 GKVPFLAVATSSAKTSPKLLDPLADNHYGQIAPKREMSQOKSEKTAFAKKDITLSL 1620
 1561 GKVPFLAVATSSAKTSPKLLDPLADNHYGQIAPKREMSQOKSEKTAFAKKDITLSL 1620
 1621 NACESNHAIAINSGOKPPEIETVNAKQSRERLCLSONPVYLKRHORETTPTLQSOOE 1680
 1621 NACESNHAIAINSGOKPPEIETVNAKQSRERLCLSONPVYLKRHORETTPTLQSOOE 1680
 1681 IDYDDTISYEMKKEEDFDIYDEDENOSPSFOKTRHIFIAAVERLMDYQSSSPHYLNR 1740
 1681 IDYDDTISYEMKKEEDFDIYDEDENOSPSFOKTRHIFIAAVERLMDYQSSSPHYLNR 1740
 1741 AQSQSVQPKRVVFOETDGSFTQPLYRGELNEHLGLGYYIABEVEDNIMVTERNOASR 1800
 1741 AQSQSVQPKRVVFOETDGSFTQPLYRGELNEHLGLGYYIABEVEDNIMVTERNOASR 1800
 1801 PYSFYSLSIYSEEDQROGAPRRKNEFKPNETYTFYEMKVOHHMAPTDEDFDCANAYFSDV 1860
 1801 PYSFYSLSIYSEEDQROGAPRRKNEFKPNETYTFYEMKVOHHMAPTDEDFDCANAYFSDV 1860
 1861 DLEKDVHSLIGPLVYCHTNTLPAARGRVYVQOEALFPFIIPEBTHSWTFENNERCRA 1920
 1861 DLEKDVHSLIGPLVYCHTNTLPAARGRVYVQOEALFPFIIPEBTHSWTFENNERCRA 1920
 1921 PCNIOHEDPTFEKENRHAINGIIMDTLGLVMAODDKIRIMYLLISGNSENHSHHSJSGH 1980
 1921 PCNIOHEDPTFEKENRHAINGIIMDTLGLVMAODDKIRIMYLLISGNSENHSHHSJSGH 1980
 1981 VFTVRKKEEYKMALYNTYPGVFETVENKLSKAGIMRVECLIGBHLAAGNSTFLVYSNKC 2040
 1981 VFTVRKKEEYKMALYNTYPGVFETVENKLSKAGIMRVECLIGBHLAAGNSTFLVYSNKC 2040
 2041 QTPPLGASGHINDFOITASQGYQMAPKLARLHYSGSINMSTRKEPSSWIKVDLAPYI 2100
 2041 QTPPLGASGHINDFOITASQGYQMAPKLARLHYSGSINMSTRKEPSSWIKVDLAPYI 2100
 2101 HGKKTGGAQKQSSLYISQFLIMYSLDGAKKQYTRNGNSTGLWVFGGNVDSGIGKHNFN 2160

```
Db 2101 HGIKIGARQKFFSSLYISOFILIMSLDGKKMTYGNSTGILWVFPGNDSSGIGHNIFN 2160
Qy 2161 PPIIARIIRLHPTHTYSTRSTLRHMLGCDLNSCNPJGAMESKAIISDAQTASSYFTNFEA 2220
Db 2161 PPIIARIIRLHPTHTYSTRSTLRHMLGCDLNSCNPJGAMESKAIISDAQTASSYFTNFEA 2220
Qy 2221 TWSPSKARLHLQGSNMRPQVNNPKEMQLQVDFQRTKRVGTGVTQGVKSILTSMTYKEFL 2280
Db 2221 TWSPSKARLHLQGSNMRPQVNNPKEMQLQVDFQRTKRVGTGVTQGVKSILTSMTYKEFL 2280
Qy 2281 ISSSDGHWMTLEFGNGKVFYFQGNDSFTPVVNSLDPELITRLRLRHQSWHVDIALRM 2340
Db 2281 ISSSDGHWMTLEFGNGKVFYFQGNDSFTPVVNSLDPELITRLRLRHQSWHVDIALRM 2340
Qy 2341 EVLGCENADLY 2351
Db 2341 EVLGCENADLY 2351

RESULT 2
FAS_MOUSE STANDARD; PRT; 2319 AA.
ID 006194;
AC 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Coagulation factor VIII precursor (Procoagulant component).
GN F8 OR CPG OR F8C.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6 X CBA; TISSUE=Liver;
RX MEDLINE=93300511; PubMed=8314577;
RA Elder B., Laskich D., Gletschler J.;
RT "Sequence of the murine factor VIII cDNA.";
RL Genomics 16:374-379(1993).
CC -1- FUNCTION: FACTOR VIII, ALONG WITH CALCIUM AND PHOSPHOLIPID, ACTS
CC AS A COFACTOR FOR FACTOR IXA WHEN IT CONVERTS FACTOR X TO THE
CC ACTIVATED FORM, FACTOR Xa.
CC -1- SUBCELLULAR LOCATION: Extracellular.
CC -1- TISSUE SPECIFICITY: FOUND IN MOST TISSUES.
CC -1- SIMILARITY: CONTAINS 3 F5/8 TYPE A DOMAINS; EACH IS COMPOSED OF
CC 2 PLASTOCYANIN-LIKE REPEATS.
CC -1- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
CC -1- SIMILARITY: STRONG, TO COAGULATION FACTOR V.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
DR EMBL: L05573; AAA37385.1; -.
DR PIR: A47004; A47004.
DR HSSP: P00431; IC8G.
DR MGD: MGI:88383; F8.
DR InterPro: IPR001117; Cu-oxidase.
DR Pfam: PF00394; Cu-oxidase; 3.
DR Pfam: PF00754; F5_F8_type_C; 2.
DR SMART: SM00231; FAS8C; 2.
DR PROSITE: PS00079; MULTICOOPER_OXIDASEL; 3.
DR PROSITE: PS01285; FAS8C_1; 2.
DR PROSITE: PS01286; FAS8C_2; 2.
DR Blood coagulation; Repeat; Plasma.
KW Signal; Glycoprotein; Sulfation.
FT SIGNAL 1 19 POTENTIAL.
```

```
FT CHAIN 20 2319
FT DOMAIN 20 349
FT DOMAIN 20 199
FT DOMAIN 207 349
FT DOMAIN 399 573
FT DOMAIN 583 730
FT DOMAIN 760 1640
FT DOMAIN 1683 2008
FT DOMAIN 1855 2008
FT DOMAIN 2009 2156
FT DOMAIN 2161 2313
FT SITE 391 392
FT SITE 750 760
FT SITE 1678 1679
FT SITE 1678 1679
FT SITE 1640 1641
FT SITE 1640 1641
FT MOD_RES 367 367
FT MOD_RES 737 737
FT MOD_RES 738 738
FT MOD_RES 742 742
FT MOD_RES 1669 1669
FT MOD_RES 1687 1687
FT DISULFID 173 199
FT DISULFID 547 573
FT DISULFID 1819 1845
FT DISULFID 2008 2156
FT DISULFID 2161 2313
FT CARBOHYD 51 61
FT CARBOHYD 233 233
FT CARBOHYD 259 259
FT CARBOHYD 423 423
FT CARBOHYD 601 601
FT CARBOHYD 880 880
FT CARBOHYD 958 958
FT CARBOHYD 1015 1015
FT CARBOHYD 1022 1022
FT CARBOHYD 1026 1026
FT CARBOHYD 1044 1044
FT CARBOHYD 1076 1076
FT CARBOHYD 1087 1087
FT CARBOHYD 1126 1136
FT CARBOHYD 1161 1161
FT CARBOHYD 1182 1192
FT CARBOHYD 1255 1255
FT CARBOHYD 1268 1268
FT CARBOHYD 1273 1273
FT CARBOHYD 1274 1274
FT CARBOHYD 1302 1302
FT CARBOHYD 1316 1316
FT CARBOHYD 1340 1340
FT CARBOHYD 1378 1378
FT CARBOHYD 1797 1797
FT CARBOHYD 2105 2105
FT SEQUENCE 2319 AA; 266148 MW; FD054DB051DB2A01 CRC64;

Query Match 71.18; Score 8835; DB 1; Length 2319;
Best Local Similarity 72.38; Pred. No. 0;
Matches 1713; Conservative 232; Mismatches 363; Indels 54; Gaps 22;

CONGLUTINATION FACTOR VIII.
F5/8 TYPE A 1.
PLASTOCYANIN-LIKE 1.
PLASTOCYANIN-LIKE 2.
F5/8 TYPE A 2.
PLASTOCYANIN-LIKE 3.
PLASTOCYANIN-LIKE 4.
B.
F5/8 TYPE A 3.
PLASTOCYANIN-LIKE 5.
PLASTOCYANIN-LIKE 6.
F5/8 TYPE C 1.
F5/8 TYPE C 2.
CLEAVAGE (BY THROMBIN) (BY SIMILARITY).
CLEAVAGE (BY THROMBIN) (BY SIMILARITY).
CLEAVAGE (BY THROMBIN) (BY SIMILARITY).
CLEAVAGE (ACTIVATION) (BY SIMILARITY).
SULFATION (BY SIMILARITY).
SULFATION (BY SIMILARITY).
SULFATION (BY SIMILARITY).
SULFATION (REQUIRED FOR VWF BINDING) (BY SIMILARITY).
SULFATION (BY SIMILARITY).
PROBABE.
PROBABE.
BY SIMILARITY.
BY SIMILARITY.
```



```

360 EEADYDDDLTDEMDVYREFDDNSPSEFIOISVAKKHPKMYVIAAEEEDMVAPLY 419
361 EE-EDIDNLDSDMDVYALDGDVSPFIQIRSVAKKHPKMYVIAAEEEDMVAAP 419
420 AADDRSYKQYLNNGPQIRGRKRYKRVMAVYDFEFTKREALQHESGILLYGEGDY 479
421 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 479
420 SPDSRSKSLYLNSGFORIGRKRYKRVMAVYDFEFTKREALQHESGILLYGEGDY 479
480 LLIIFKNOASRPYNIYPIGIDVAPRLYRRLPGVYKHLKDFILPGEIYKMYVYEDG 539
480 LLIIFKNOASRPYNIYPIGIDVAPRLYRRLPGVYKHLKDFILPGEIYKMYVYEDG 539
480 LLIIFKNOASRPYNIYPIGIDVAPRLYRRLPGVYKHLKDFILPGEIYKMYVYEDG 539
540 PTKSDPRLCTRYSSVYVMEEDLASGLIGPLICRYESYDOGNOIMSDRNYLSEVD 599
540 PTKSDPRLCTRYSSVYVMEEDLASGLIGPLICRYESYDOGNOIMSDRNYLSEVD 599
540 PTKSDPRLCTRYSSVYVMEEDLASGLIGPLICRYESYDOGNOIMSDRNYLSEVD 599
600 ENRSWYLTENIQEFLPNPAGVOLDPEFOASNIMHSINGYVDSIQLSVCLHEVAYTL 659
600 ENRSWYLTENIQEFLPNPAGVOLDPEFOASNIMHSINGYVDSIQLSVCLHEVAYTL 659
600 ENRSWYLTENIQEFLPNPAGVOLDPEFOASNIMHSINGYVDSIQLSVCLHEVAYTL 659
660 SIGAOTDFLSVPSGYTFKHKMYEDLTLFPFSGETVMSKENGMLIIGCHNSDFRR 719
660 SIGAOTDFLSVPSGYTFKHKMYEDLTLFPFSGETVMSKENGMLIIGCHNSDFRR 719
660 SIGAOTDFLSVPSGYTFKHKMYEDLTLFPFSGETVMSKENGMLIIGCHNSDFRR 719
720 GMTALLKVSCKDNTEDEYEDISATLLSKNNAIEPRSPSONSRPSTROKOFNAT 779
720 GMTALLKVSCKDNTEDEYEDISATLLSKNNAIEPRSPSONSRPSTROKOFNAT 779
720 GMTALLKVSCKDNTEDEYEDISATLLSKNNAIEPRSPSONSRPSTROKOFNAT 779
760 IPENDIEKTPWPAHHTPMKIQONVSSDLMMLROSPTPHGLSLSDQEKAYTESDDP 839
760 IPENDIEKTPWPAHHTPMKIQONVSSDLMMLROSPTPHGLSLSDQEKAYTESDDP 839
760 IPENDIEKTPWPAHHTPMKIQONVSSDLMMLROSPTPHGLSLSDQEKAYTESDDP 839
760 IPENDIEKTPWPAHHTPMKIQONVSSDLMMLROSPTPHGLSLSDQEKAYTESDDP 839
840 SPGAIDSNNSLSEMTFHPRLHSGDMVTFPESGLOLRNEKLTGTAATELKIDKIDFVSS 899
840 SPGAIDSNNSLSEMTFHPRLHSGDMVTFPESGLOLRNEKLTGTAATELKIDKIDFVSS 899
840 SPGAIDSNNSLSEMTFHPRLHSGDMVTFPESGLOLRNEKLTGTAATELKIDKIDFVSS 899
837 LPGAERNTNPSAAALRRLHLSAERYLPEP-----EK-----ELKIDSDMS 882
900 TSNMLIS--TIPADNIAAGTNTSLSGPPSPYHYDOLDTLFEGKSSPLTESGGLSL 957
900 TSNMLIS--TIPADNIAAGTNTSLSGPPSPYHYDOLDTLFEGKSSPLTESGGLSL 957
900 TSNMLIS--TIPADNIAAGTNTSLSGPPSPYHYDOLDTLFEGKSSPLTESGGLSL 957
883 SSDLLKTPSPITPSDITLSEETERTSHLPPHYPVNSRQALIVLCKNSHRTIGAVPLS 942
958 SEENNDKLLSEGLMSOESSGKNGVSTESGFLPKGRARHAPALLTDNALFVYSIL 1017
958 SEENNDKLLSEGLMSOESSGKNGVSTESGFLPKGRARHAPALLTDNALFVYSIL 1017
958 SEENNDKLLSEGLMSOESSGKNGVSTESGFLPKGRARHAPALLTDNALFVYSIL 1017
943 TEEDH-----SSIGENYVSTESDGIFFERKARHAPALLTDNALFVYSIL 989
1018 KTKNTSNNSATNKKYHIDGPSLLIENSPYQWNLISDTEFRKVPPLIHDMMDKNATA 1077
1018 KTKNTSNNSATNKKYHIDGPSLLIENSPYQWNLISDTEFRKVPPLIHDMMDKNATA 1077
1018 KTKNTSNNSATNKKYHIDGPSLLIENSPYQWNLISDTEFRKVPPLIHDMMDKNATA 1077
990 KTKNARYLTKNRIHIDDAALLTENRASA-----TFEDKWTJA 1028
1078 LRLNHSNKTSSKNKEMVQOKKEGPIPPDAQNDMSFFKMLFEPESARWIOETHGKNSI 1137
1078 LRLNHSNKTSSKNKEMVQOKKEGPIPPDAQNDMSFFKMLFEPESARWIOETHGKNSI 1137
1078 LRLNHSNKTSSKNKEMVQOKKEGPIPPDAQNDMSFFKMLFEPESARWIOETHGKNSI 1137
1029 SGLNHYEN-----WIKGPIKKNPL 1047
1138 NSGQSPKQVSLGPEKSEVCONFLSEKKNVYVKGFTXDVGLKEMVPPSSNLFITN 1197
1138 NSGQSPKQVSLGPEKSEVCONFLSEKKNVYVKGFTXDVGLKEMVPPSSNLFITN 1197
1138 NSGQSPKQVSLGPEKSEVCONFLSEKKNVYVKGFTXDVGLKEMVPPSSNLFITN 1197
1048 SBERGPELTLSSGSGKSVKSGGGRIRAVVEEELSGK--KEMMLNLSLFTFLN 1104
1198 LDIENHNTNOKKLOEIEKKEKTLIOENVYLOIHYTQKTKNKKMLFLSLRQVWG 1257
1198 LDIENHNTNOKKLOEIEKKEKTLIOENVYLOIHYTQKTKNKKMLFLSLRQVWG 1257
1198 LDIENHNTNOKKLOEIEKKEKTLIOENVYLOIHYTQKTKNKKMLFLSLRQVWG 1257
1105 SADVOGNDTHSGCKREKREKREKREKREKREKREKREKREKREKREKREKREK 1164
1258 SYDAVAPVLODFRSLNDSTNFTKKNHAKHSKKEEENLEGLNOYKOIYEVACTYRIS 1317
1258 SYDAVAPVLODFRSLNDSTNFTKKNHAKHSKKEEENLEGLNOYKOIYEVACTYRIS 1317
1258 SYDAVAPVLODFRSLNDSTNFTKKNHAKHSKKEEENLEGLNOYKOIYEVACTYRIS 1317
1165 FDGGSRAVPVODRSRLNDSAEARETHLAHSAIAREAPLEAGNMT----- 1210
1318 PNTSQONFYQSRKRALOKFLPEETELERIIIVDTSTQSKMKHILPSTLTQIDYN 1377
1318 PNTSQONFYQSRKRALOKFLPEETELERIIIVDTSTQSKMKHILPSTLTQIDYN 1377
1318 PNTSQONFYQSRKRALOKFLPEETELERIIIVDTSTQSKMKHILPSTLTQIDYN 1377
1211 -GPGRSRAVPRVKQSLQIRLPLEIKPERGVVLANVSTRMS----- 1252
1378 EKEGALTQSPISDCILTRSHSIPQANRSLPLAKVSSPFSIRPIYTLVTLQODNSSHLP 1437
1253 ----- 1252
1438 ASYRKDSGVQESSHPLQGAKKNNLSAILTELENTGDQREVGSLGTSATNSVYTKKVENT 1497

```

```

1253 -----SSSIDGAKNNLSPLTLEMAAGGOKTSALGSAAGPLASKEKA 1301
1253 -----SSSIDGAKNNLSPLTLEMAAGGOKTSALGSAAGPLASKEKA 1301
1253 -----SSSIDGAKNNLSPLTLEMAAGGOKTSALGSAAGPLASKEKA 1301
1498 VLKRPDLPTSGKVELLPKRYHIOKDLPEFTENSSPGHLDLVEGSLQGTGAIKNEA 1557
1498 VLKRPDLPTSGKVELLPKRYHIOKDLPEFTENSSPGHLDLVEGSLQGTGAIKNEA 1557
1498 VLKRPDLPTSGKVELLPKRYHIOKDLPEFTENSSPGHLDLVEGSLQGTGAIKNEA 1557
1302 VLSAGLSASGAKAFEPKRVNHEEDLPOKTSNYSCHAGSLGGEIPLQRTGPGVNLKV 1361
1558 NRKGPFLRVATESSAKTPEKLLDPLANDNHYQIPIREEMKSOEKSEPKTAFKRTDI 1617
1558 NRKGPFLRVATESSAKTPEKLLDPLANDNHYQIPIREEMKSOEKSEPKTAFKRTDI 1617
1558 NRKGPFLRVATESSAKTPEKLLDPLANDNHYQIPIREEMKSOEKSEPKTAFKRTDI 1617
1362 NRKG-----RTPSKLLGP-----PMK-EMSELEKSPRYALATKDI 1398
1618 -LSLNNCSNHAIAINQONKPEIETAKOGRERKCSQNPVYKRRORRTTQOS 1676
1618 -LSLNNCSNHAIAINQONKPEIETAKOGRERKCSQNPVYKRRORRTTQOS 1676
1618 -LSLNNCSNHAIAINQONKPEIETAKOGRERKCSQNPVYKRRORRTTQOS 1676
1399 SLPLDHESSNHAIAINQONKPEIETAKOGRERKCSQNPVYKRRORRTTQOS 1458
1677 DQEBIDYDFTISVENKEDPDIDEDENOSPPSPFOKTRHAFIAVERLMDYGSSSPHV 1736
1677 DQEBIDYDFTISVENKEDPDIDEDENOSPPSPFOKTRHAFIAVERLMDYGSSSPHV 1736
1677 DQEBIDYDFTISVENKEDPDIDEDENOSPPSPFOKTRHAFIAVERLMDYGSSSPHV 1736
1459 EEDKMDYDFTISVENKEDPDIDEDENOSPPSPFOKTRHAFIAVERLMDYGSSSPHV 1518
1737 LNRNAGSGVPOFKKRVYQOEFTEGDFTOPLRYGELNENHGLGPIYRAVEONIMVTERN 1796
1737 LNRNAGSGVPOFKKRVYQOEFTEGDFTOPLRYGELNENHGLGPIYRAVEONIMVTERN 1796
1737 LNRNAGSGVPOFKKRVYQOEFTEGDFTOPLRYGELNENHGLGPIYRAVEONIMVTERN 1796
1519 LNRNAGSGVPOFKKRVYQOEFTEGDFTOPLRYGELNENHGLGPIYRAVEONIMVTERN 1578
1797 QASRPYSFYSLSIYEEPOGAGAPKRNPFYKRNKRYFMKVQHMAPTKDEPCKANAY 1856
1797 QASRPYSFYSLSIYEEPOGAGAPKRNPFYKRNKRYFMKVQHMAPTKDEPCKANAY 1856
1797 QASRPYSFYSLSIYEEPOGAGAPKRNPFYKRNKRYFMKVQHMAPTKDEPCKANAY 1856
1579 QASRPYSFYSLSIYEEPOGAGAPKRNPFYKRNKRYFMKVQHMAPTKDEPCKANAY 1638
1657 FSDYDLKRYHSGGLIGPLICRYESYDOGNOIMSDRNYLSEVD 1916
1657 FSDYDLKRYHSGGLIGPLICRYESYDOGNOIMSDRNYLSEVD 1916
1657 FSDYDLKRYHSGGLIGPLICRYESYDOGNOIMSDRNYLSEVD 1916
1659 FSDYDLKRYHSGGLIGPLICRYESYDOGNOIMSDRNYLSEVD 1698
1917 NCRAPCNIOMEDPTEKREYRERHAINGYIMOTLPGVMAQDORIMWYLLSGNSNENHSH 1976
1917 NCRAPCNIOMEDPTEKREYRERHAINGYIMOTLPGVMAQDORIMWYLLSGNSNENHSH 1976
1917 NCRAPCNIOMEDPTEKREYRERHAINGYIMOTLPGVMAQDORIMWYLLSGNSNENHSH 1976
1699 NCRAPCNIOMEDPTEKREYRERHAINGYIMOTLPGVMAQDORIMWYLLSGNSNENHSH 1758
1977 FSGHYFTVKKKEEYKALYNYLPSEVEYEMKPSKAGIMRECELGHEHLAGSTLEFY 2036
1977 FSGHYFTVKKKEEYKALYNYLPSEVEYEMKPSKAGIMRECELGHEHLAGSTLEFY 2036
1977 FSGHYFTVKKKEEYKALYNYLPSEVEYEMKPSKAGIMRECELGHEHLAGSTLEFY 2036
1759 FSGHYFTVKKKEEYKALYNYLPSEVEYEMKPSKAGIMRECELGHEHLAGSTLEFY 1818
2037 SNKOPPLGMAAGHIDROITASQOYONAPKILAHYSSGINSNMSTRKEPFSYIVYDLA 2096
2037 SNKOPPLGMAAGHIDROITASQOYONAPKILAHYSSGINSNMSTRKEPFSYIVYDLA 2096
2037 SNKOPPLGMAAGHIDROITASQOYONAPKILAHYSSGINSNMSTRKEPFSYIVYDLA 2096
1819 SKBOAPLMSAGRIROITASQOYONAPKILAHYSSGINSNMSTRKEPFSYIVYDLA 1878
2057 PMIHGKTQGAROKFSSLYISOFITMSLDGKNMOTYRGNSTGLTWFFGVNDSSGIRK 2156
2057 PMIHGKTQGAROKFSSLYISOFITMSLDGKNMOTYRGNSTGLTWFFGVNDSSGIRK 2156
2057 PMIHGKTQGAROKFSSLYISOFITMSLDGKNMOTYRGNSTGLTWFFGVNDSSGIRK 2156
1879 PMIHGKTQGAROKFSSLYISOFITMSLDGKNMOTYRGNSTGLTWFFGVNDSSGIRK 1938
2157 NIFNPPIIARYIRLPHYRSIRSLRMEIAGCDLNSCSMPGMEKSAISDAQITASSYFT 2216
2157 NIFNPPIIARYIRLPHYRSIRSLRMEIAGCDLNSCSMPGMEKSAISDAQITASSYFT 2216
2157 NIFNPPIIARYIRLPHYRSIRSLRMEIAGCDLNSCSMPGMEKSAISDAQITASSYFT 2216
1939 NIFNPPIIARYIRLPHYRSIRSLRMEIAGCDLNSCSMPGMEKSAISDAQITASSYFT 1998
2217 NMFATSPSKARLHIOGRSNMROVNPKEMLQVDFOKTAKVYGVTOGVKSLTSMY 2276
2217 NMFATSPSKARLHIOGRSNMROVNPKEMLQVDFOKTAKVYGVTOGVKSLTSMY 2276
2217 NMFATSPSKARLHIOGRSNMROVNPKEMLQVDFOKTAKVYGVTOGVKSLTSMY 2276
1999 NIFATWSPQARLHIOGRSNMROVNPKEMLQVDFOKTAKVYGVTOGVKSLTSMY 2058
2277 KEPLISSODGHOMLTFONKRYVFGONDSTPRVNSLPLTLYLHINQSMYNOI 2336
2277 KEPLISSODGHOMLTFONKRYVFGONDSTPRVNSLPLTLYLHINQSMYNOI 2336
2277 KEPLISSODGHOMLTFONKRYVFGONDSTPRVNSLPLTLYLHINQSMYNOI 2336
2059 KEPLISSODGHOMLTFONKRYVFGONDSTPRVNSLPLTLYLHINQSMYNOI 2118
2337 ALRREVLGCEADLY 2351
2337 ALRREVLGCEADLY 2351
2337 ALRREVLGCEADLY 2351
2119 ALRREVLGCEADLY 2133

```

```

RESULT 4
FAS_HUMAN
ID FAS_HUMAN STANDARD: PRT: 2224 AA.
AC P12259: 014285:
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Coagulation factor V precursor (Activated protein C cofactor).
GN F5.

```

OS Homo sapiens (Human).
 OC Eumaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 NCBI_Taxid:9606;
 RN [1]
 RN SEQUENCE FROM N.A.
 RX MEDLINE-92232668; PubMed-1567832;
 RA Cripe L.D., Moore K.D., Kane W.H.;
 RT "Structure of the gene for human coagulation factor V.";
 RL Biochemistry 31:3777-3783(1992).
 RN [2]
 RN SEQUENCE FROM N.A.
 RX MEDLINE-87260886; PubMed-3110773;
 RA Jenny R.J., Pittman D.D., Toole J.J., Kriz R.W., Aldape R.A.,
 RA Hewick R.M., Kaufman R.J., Mann K.G.;
 RT "Complete cDNA and derived amino acid sequence of human factor V.";
 RL Proc. Natl. Acad. Sci. U.S.A. 84:4846-4850(1987).
 RN [3]
 RN SEQUENCE OF 1-1600 FROM N.A.
 RX MEDLINE-88107560; PubMed-2827731;
 RA Kane W.H., Ichinose A., Hagen F.S., Davie E.W.;
 RT "Cloning of cDNAs coding for the heavy chain region and connecting
 RT region of human factor V, a blood coagulation factor with four types
 RT of internal repeats.";
 RL Biochemistry 26:6508-6514(1987).
 RN [4]
 RN SEQUENCE OF 1188-1215 AND 1315-2224 FROM N.A.
 RX MEDLINE-86313665; PubMed-3092220;
 RA Kane W.H., Davie E.W.;
 RT "Cloning of a cDNA coding for human factor V, a blood coagulation
 RT factor homologous to factor VIII and ceruloplasmin.";
 RL Proc. Natl. Acad. Sci. U.S.A. 83:6800-6804(1986).
 RN [5]
 RN PARTIAL SEQUENCE FROM N.A.
 RP TISSUE-Fibroblast;
 RX MEDLINE-93203619; PubMed-8454869;
 RA Shen N.L.L., Fan S.-T., Pyatt J., Graff R., Lapolla R.J.,
 RA Edgington T.S.;
 RT "The serine protease cofactor factor V is synthesized by
 RT lymphocytes.";
 RL J. Immunol. 150:2992-3001(1993).
 RN [6]
 RN SULEFATON.
 RP MEDLINE-94264012; PubMed-8204629;
 RX Pittman D.D., Tomkinson K.N., Michnick D., Seligsohn U.,
 RA Kaufman R.J.;
 RT Posttranslational sulfation of factor V is required for efficient
 RT thrombin cleavage and activation and for full prococagulant activity.";
 RL Biochemistry 33:6952-6959(1994).
 RN [7]
 RN SULEFATON.
 RP MEDLINE-90366699; PubMed-2168225;
 RX Horton G.L.;
 RA "Sulfation of tyrosine residues in coagulation factor V.";
 RL Blood 76:946-952(1990).
 RN [8]
 RN X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 2065-2224.
 RX MEDLINE-20052169; PubMed-10586886;
 RA Macedo-Ribeiro S., Bode W., Huber R., Quin-Alen M.A., Kim S.W.,
 RA Ortel T.L., Bourenkov G.P., Bartunik H.D., Stubbs M.T., Kane W.H.,
 RA Fuentes-Prior P.;
 RT "Crystal structures of the membrane-binding C2 domain of human
 RT coagulation factor V.";
 RL Nature 402:434-439(1999).
 RN [9]
 RN VARIANT APCR GIN-534.
 RX MEDLINE-94217810; PubMed-8164741;
 RA Bertina R.M., Koelman B.P.C., Koster T., Rosendaal F.R.,
 RA Dirven R.J., de Ronde H., van der Velden P.A., Reitsma P.H.;
 RT "Mutation in blood coagulation factor V associated with resistance to
 RT activated protein C.";
 RL Nature 369:64-67(1994).
 RN [10]

RP VARIANT APCR GUY-334, AND VARIANT LYS-513.
 RA Chan W.P., Lee C.K., Kwong Y.L., Lam C.K., Liang R.;
 RT "A novel mutation of Arg306 of factor V gene in Hong Kong Chinese.";
 RL Blood 91:1135-1139(1998).
 RN [11]
 RN VARIANT APCR THR-334.
 RA Williamson D., Brown K., Luddington R., Baglin C., Baglin T.;
 RT "Factor V Cambridge: a new mutation (Arg306-to-Thr) associated with
 RT resistance to activated protein C.";
 RL Blood 91:1140-1144(1998).
 CC -1- FUNCTION: COAGULATION FACTOR V IS A COFACTOR THAT PARTICIPATES
 CC WITH FACTOR XA TO ACTIVATE PROTHROMBIN TO THROMBIN.
 CC -1- SUBUNIT: Factor Va is composed of a heavy chain and a light
 CC chain, noncovalently bound. The interaction between the two chains
 CC is calcium-dependent.
 CC -1- DOMAIN: DOMAIN B CONTAINS 35 X 9 AA TANDEM REPEATS, AND 2 X 17 AA
 CC REPEATS.
 CC -1- PTM: Thrombin activates factor V proteolytically to the active
 CC cofactor, factor Va (formation of a heavy chain at the N-
 CC terminus and a light chain at the C-terminus).
 CC -1- PTM: SULFATION IS REQUIRED FOR EFFICIENT THROMBIN CLEAVAGE AND
 CC ACTIVATION AND FOR FULL PROCOCAGULANT ACTIVITY.
 CC -1- DISEASE: OMEN PARAHEMOPHILIA, AN HEMORRHAGIC DIAPYRESIS, IS DUE
 CC TO A DEFICIENCY OF FACTOR V. OTHER DEFECTS IN F5 RESULTS IN A
 CC FORM OF THROMBOPHILIA KNOWN AS APC RESISTANCE (APCR). THE APCR
 CC MUTATION IS FOUND IN ABOUT 5% OF THE POPULATION WHICH SUGGEST THAT
 CC A SLIGHT THROMBOTIC TENDENCY MAY CONFER SOME ADVANTAGE IN FETAL
 CC IMPLANTATION.
 CC -1- SIMILARITY: CONTAINS 3 F5/8 TYPE A DOMAINS; EACH IS COMPOSED OF
 CC 2 PLASTOCYANIN-LIKE REPEATS.
 CC -1- SIMILARITY: CONTRAINS 2 F5/8 TYPE C DOMAINS.
 CC -1- SIMILARITY: STRONG, TO COAGULATION FACTOR VIII.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation.
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: L32779; AAB59401.1; JOINED.
 DR EMBL: L32755; AAB59401.1; JOINED.
 DR EMBL: L32756; AAB59401.1; JOINED.
 DR EMBL: L32757; AAB59401.1; JOINED.
 DR EMBL: L32758; AAB59401.1; JOINED.
 DR EMBL: L32759; AAB59401.1; JOINED.
 DR EMBL: L32760; AAB59401.1; JOINED.
 DR EMBL: L32761; AAB59401.1; JOINED.
 DR EMBL: L32762; AAB59401.1; JOINED.
 DR EMBL: L32763; AAB59401.1; JOINED.
 DR EMBL: L32764; AAB59401.1; JOINED.
 DR EMBL: L32765; AAB59401.1; JOINED.
 DR EMBL: L32766; AAB59401.1; JOINED.
 DR EMBL: L32767; AAB59401.1; JOINED.
 DR EMBL: L32768; AAB59401.1; JOINED.
 DR EMBL: L32769; AAB59401.1; JOINED.
 DR EMBL: L32770; AAB59401.1; JOINED.
 DR EMBL: L32771; AAB59401.1; JOINED.
 DR EMBL: L32772; AAB59401.1; JOINED.
 DR EMBL: L32773; AAB59401.1; JOINED.
 DR EMBL: L32774; AAB59401.1; JOINED.
 DR EMBL: L32775; AAB59401.1; JOINED.
 DR EMBL: L32776; AAB59401.1; JOINED.
 DR EMBL: L32777; AAB59401.1; JOINED.
 DR EMBL: L32778; AAB59401.1; JOINED.
 DR EMBL: L32779; AAB59401.1; JOINED.
 DR EMBL: M16967; AAB52424.1; JOINED.
 DR EMBL: M14353; AAB59532.1; JOINED.
 DR PIR: A25897; A25897.
 DR PIR: A28028; A28028.
 DR PDB: 1CZS; 26-NOV-99.
 DR PDB: 1CZT; 26-NOV-99.
 DR PDB: 1CZV; 26-NOV-99.

DR Genew: HGNC:3542; F5.
DR MIM: 227400; -.
DR MIM: 134400; -.
DR MIM: 188055; -.
DR MIM: 227310; -.
DR InterPro: IPR001117; Cu-oxidase.
DR InterPro: IPR000421; FA58_C.
DR Pfam: PF00394; Cu-oxidase; 3.
DR Pfam: PF00754; F5_F8_type_C; 2.
DR SMART: SM00231; FA58C; 2.
DR PROSITE: PS00079; MULTICOPPER_OXIDASE1; 2.
DR PROSITE: PS01285; FA58C_1; 2.
DR PROSITE: PS01286; FA58C_2; 2.
DR Blood coagulation; Glycoprotein; Sulfation; Calcium; Signal; Zymogen;
KM Repeat; Polymorphism; Disease mutation; Thrombophilia; 3d-structure.
FT SIGNAL 1 28
FT CHAIN 29 2224
FT CHAIN 29 737
FT PEPTIDE 738 1573
FT CHAIN 1574 2224
FT DOMAIN 30 329
FT DOMAIN 203 329
FT DOMAIN 348 684
FT DOMAIN 536 684
FT DOMAIN 692 1573
FT DOMAIN 895 928
FT REPEAT 912 928
FT REPEAT 912 1148
FT SIMILAR 1135 1148
FT DOMAIN 1185 1501
FT REPEAT 1185 1193
FT REPEAT 1194 1202
FT REPEAT 1203 1211
FT REPEAT 1212 1220
FT REPEAT 1221 1229
FT REPEAT 1230 1238
FT REPEAT 1239 1247
FT REPEAT 1248 1256
FT REPEAT 1257 1265
FT REPEAT 1266 1274
FT REPEAT 1275 1283
FT REPEAT 1284 1292
FT REPEAT 1293 1301
FT REPEAT 1302 1310
Query Match 22.7%; Score 2816.5; DB 1; Length 2224;
Best Local Similarity 30.5%; Pred. No. 1.5e-138;
Matches 768; Conservative 387; Mismatches 853; Indels 507; Gaps 77;
22 RRYVIGAVELSDMYQSDGLGELPVDAPPPRVKSPFNTSVY-YKKTLPVEETDHLFNI 80
32 RQFYVAGISMSYRE-----PTNSLMLSTSKTKYVREVEPT-FRK 75
81 AKRPPPMGLGPTIOAEYVTFVITLKNASHPVSLAAGVSYWKASEGAEIDDOTSOR 140
76 EKPGSTISGLGPTLAEVGLIRVHKRADRPISIHQGIKRSKISEGSLDITFPFA 135
141 EKEDKVPFGSHTYVQYLKENGPMASDPLCTYSLSHVDVLVDNLGSLIGALLVCEB 200
136 EKMDAVALPGREYTYEMSISSDSGPTDDPPCLTHYISHENLIDFNLSGLIPLLICK 195
201 GSLAKKETO-TLHK-TILLFVFPDGKSWHSEFKNLMDRMAASARWPMKFTVNGYN 258
196 GTLEGGTOKTTPKQVYLFAVFDSEKSSOSS-----LMTYVNGYN 239
259 RSLPGLIGCHRSKSVWVIGKTPPEVHSTFLSGHTPLVRNHRQASLEISPTPLRQTL 318
240 GTPMDIVVCAHDHISHLGMSGSELPSTHFNGOVLQDNHHRKVSATLVLASASTTANMT 299
319 LMDLGQFLFLCHISHQHDGMEAVYKVDSCEPPEPOLMRKNNEADYDDDLTDEMDYVR 378

DB 300 VQPEGKWIISLTPKHLQAGQAVYIDIKKCPKTRNLKTTRE----- 342
DB 379 EDDNSPSFIQISVAKKRPVWVYIAAEEDMDYAPLVLPADDSRYKSOYLANNGPQRI 438
DB 343 -----ORRHKKRWEYVIAAEVYIMVYAVPIVPMKKDRYSQHLDFNSQI 387
DB 439 GRKKKRVMAVYDEFTKREAI-----QHESSGILGLYGEVGTLLIFKKNASPRVY 495
DB 388 GKHKKVMYQYEDSEF-TKHVNPMMKIDGLIPITIRAOVRTLIVKKNASPRYSY 446
DB 496 PHGT-----DVRPLSRRLPRGVHLDLPFLIGLEFYKKTVYVEGPKSPDLT 549
DB 447 PHGYTSPVEDEVNSFT-----SGRNMTIRAVQGEYTYKNNILEPEPTEADQCLT 502
DB 550 RYSSVFNMRDLASGLIGPLLCYSEVDQGNQINSKRNILFVDEKRSYLTEN 609
DB 503 RPYSSVDIMDLASGLIGLLICKRSIDRGIQRAADIEQAVAVDEKRSYLTEDN 562
DB 610 IORFLPNAGVQLEDEPEFOASNIMHSINGVYDSL-QSYCLHEVAYWYLSIGAQTDEL 668
DB 563 INFCEPNDEVRDPPKFPESNININSTINGVSEITVIGFCFDVQWHPSCVGTQNEIL 622
DB 669 SVFSGYTKRMVYEDTLTPFSGEYVFNEMKVPGLWIGCHNSDPFRMGATLAYS 728
DB 623 THFTGHSPIYKRRHEDTLTPMRGESYTVMDVGVMTLSNNSPRSKTRLEFDY 682
DB 729 SCQKNMGDYEDSYEDT-----SAYLSK--NNAIRP-----RSF 761
DB 683 KC--IPDDEDSYETIEPPSTVMAKRMHRLPEDESDADYQORLAAALGISPP 739
DB 762 SONSRRPSTROKFNATTP-ENDIEKTPWNAHTPMKIQNVSSDMLLRQSPTH 820
DB 740 RNSSL--NOEEEFMLPALENGTE-----FVSSNTDIIVGSYSSPS 781
DB 821 GLS-----LSDLQEA--KYETSDPSPAIDSNLSLSEMT--HFRPOLHSGDMFT 869
DB 782 NISKFTVNNLAPQAPSHQOATTAQSPILRLIGKNSVLSNSTAHSHP--YSEDIED 838
DB 870 P-----ESGLOLRNKKOTTAETLKKLDKQKYSNNLSITSDMLA----- 914
DB 839 PLOPDVYTGRL-----LSLNGEKKOEBAKKGKRVKROAKKHPMSMKLLA 887
DB 915 --AGDNTSSLGPES-MPVHYDSOLDTLTPGKSSPPLTESGSLISEENDSKILL--- 967
DB 888 HKVGRHLSDTGSPPGMWEDLPDSTGSPSRMRPMKDPSPDLLLKQSNKSKILVGRW 947
DB 968 -----ESG-----LMNSQESS--NGK-----NVSSTESGRLEFKGRAHG 999
DB 948 HLASERGSYEIIQDTEDTAVNNMLISQNASRANGESTPLANKPGKOSCHPFRVRHK 1007
DB 1000 PALITRD--NALFKYSISLKTNTKSNSATNRKTHIDGSLLENPSWQNIIESDTE 1057
DB 1008 SLOVRDGGKSLKSOPLITRKKKKEKHTH-----APLSRTPHPISEAVN 1057
DB 1058 FKRYTPLIHDRMLMDKNATYALRLHMSNKTTSKKNKMWQOKKGGPIPPDAQN--PDMSE 1115
DB 1058 TFSERRLKSLVL-----HKSNEIS-----LPDNLQOTLPSDF 1091
DB 1116 FKMLFPEGARWIORTHGKNSLN--SGGSPKQOLVSLGPEKSYGVQNF-LSEKNPVYVGK 1173
DB 1092 GWIASLPD-----HNQNSNDTQASCPGLQVGPPEE-HYGFPPIDPDQDQ--- 1138
DB 1174 GFTKDVG-----LKKMV-FPSRNLPLTNDLHNHNHNEKKI-----QEELEKKE 1221
DB 1139 -HSTSPSRHSSPPLSKMLEYDSKHSKSPDTQISQMSSESHEWQVVISIPDLSOYTLSP 1197
DB 1222 TLQIENVVLPQIHYTYGKKNKMLFLSTRONEGSDGAYAVLQDFRSLDSTNRT- 1280
DB 1198 ELQOTNLSPDLSHTLSPDLQIURL-----SPALQMPISPDLSHTTL 1240
DB 1281 ---KRTAHSKKGEEENLIGLQNOTKOIYERKYACTRISPSNTSQQNFYVORSKRLAQF 1337

Dd	1241	SPDLSHTT-----LSLDSLQ-----TNLSLSQTNL-----SPALGO-12733
Qy	1338	RLPELELEKRIIYDPTSTQMSKMMHLLP-TLQIDYENKRGKAIQTSPSLDCL---13933
Dd	1274	-MPLSPDLSTHTYLSLDFSGTNLSPELSHMLLSPELSQTNLSP-----ALGOMP-LSPLDST1328
Qy	1394	TRSHSTIOANRSLPLPIAKYSSFPSPRIYTLRTVLFQDNSSHLPAASTRKDDSCVQSESHF1453
Dd	1329	TLUSDPSQTNLSP-ELISQTNLSPLALGOMPUS-----PDSHTYLSLDSQT---1373
Qy	1454	LOGAKNNLSLALITLLEMTGDQREYSGLSGATNSVYTKYKENTVLP--KPLDPTSGK1510
Dd	1374	-----NLSPELSQTNLSPLDSEMPPLFALSOIPLTPPLDQMLSLDLETLSPNGO1426
Qy	1511	VELLPKY-----HYORDLEPTETNSGSPGHLDY-----EGSLDGTGALAKNME1556
Dd	1427	MSLSPDLSQLVTLSPDLSDPTTLLPDLISOISPPPLDQIEFSSSSQSILD-----EFNE1480
Qy	1557	ANRGRKPYPLRYVTESSATP-----SKLDPLAM-----DHHGTQIPIKEEKSOE1603
Dd	1481	S-----PPYDLOGMPSPSSPLINDFLSKSEFPIYVGLSKDGDYIEIHPREVQSE1539
Qy	1604	KSPKETAFFKKKDDILSLANCSNNAIAINEGONKEIETVMAQGRBELCQNPVLX1666
Dd	1536	-----1533
Qy	1664	RHORETRTLQSDQBEID--YDDTISVEKKEDFDIYEDENOS--PNSFOKTRHYE1718
Dd	1536	-----DDYAEIDYPPYDPPKTYVRNINSSHDPPNIAAYLNSNNGNRNY1583
Qy	1719	IAAVERLMDYGMSSPHVLRNR--AOGSGVPO--FKRYFOELTDSFTQPLRGELNE1773
Dd	1584	IAAEISMDY-----SEFYORETLEDSDIDPELTQYTKKVFYKLDSTFYKRPDSREYE1639
Qy	1774	HGLGLGPTRAEVEDNIMVTFPNQASRYSFSSLSLSE-----EDQOGAEGRKMF1825
Dd	1640	HGLIGLPIRAEVDYDIOYFRKNLASRYSGLHAGLSSEKSGEKTVEDDSPEMFKEDNA1699
Qy	1826	VBPNETKTYKQYQHMAATKDEPCDCAKMAFYDNLKEDYHSGLIGLPLVYCHNTLNA1889
Dd	1700	VOPNSITYWHATBESGSPESGACRMAKATYSAVNPENDHSLIGLPLCQKGLIHD1755
Qy	1886	HGRQYVTOERPALFTTIDETKSWTFETMENERCBAQCIONMEDPTFEKNTFPAIINGYM1945
Dd	1760	SNMPDMEFVLLFEMFDKRSWYEEKSSNR-----LTSSEKSSHEFPAIINGMY1813
Qy	1946	DTLPGLVAAODORIRYLLLSMGSNENIHSHFSGHVFYVRKKEEYKMALYNLYPVGEVY2005
Dd	1814	-SLPELKATYEDENWRRLHLNLTIGSSODIHVHFHQGLLENKRGKQHLQVWPLLPJGSKFL1872
Qy	2006	BWLPKAGIRVAVCLLIGELHLAGMSTLEPLVYNNCQTPGLMAAGHIRDFOITASQVQOM2065
Dd	1873	EKKASRKPQWMLLTVEYGENORAGQOTPLIMDRCHRMPGLSTGIILSDQIKASEFLGY1933
Qy	2066	APKRLARLHYSGINANSTRE--PRS--WXYDLLAPMIGHITOGAROKRSTLYISO2113
Dd	1933	BERLARLNNGSYNNANSYKLAERASRPAQYDMQKEVITIGIOQAKHYLKSCTYTB1999
Qy	2120	FLIYSLDGKKWQTYTGNSTGTLMYFGSNDSSGIRKHNFPPIARTIARLHPTHTSIS2177
Dd	1993	FYVAISSQIMQJFEGNSTRVNMTFNGNSDASTIKENOFPDIYARIKIRLSPRAYNRP2057
Qy	2180	TRMLIMCCDLNSCSPMLMESKATSDMQOITASSYFTNNFAT-NSPSAKRILHLOGRSM2236
Dd	2053	TLRELGCEYNGCSPPLGEMNGKJENKOIYASSFKSNMGQYMPFPARILNAGRGVNM2111
Qy	2239	BOYVNNPEMYOVDQFKTKYTGVTGQGVSKLSTMYVEPLISSQDGHQWTFEPQNK2299
Dd	2113	QAKANNKQWLEIDLKIKITAITITQCKRSLSEMYVSYTIHNSBQGVKMYRLAKSS2177
Qy	2299	V--KYFOGNODSTFPVYNSLDPPLTLRLKRIHQPSVYHOLAIRMEYLCCGKODLY2351
Dd	2173	MYDKTEBENTTKGHYKFNFPNPLISIRIATVPTWMOGSTRIRLEJFGC---DIY2224

RESULT 5		
FA5_PIG	STANDARD;	PRT: 2258 AA.
AC	09GLP1:	
DR	16-OCT-2001 (Rel. 40, Created)	
DR	16-OCT-2001 (Rel. 40, Last sequence update)	
DR	15-JUN-2002 (Rel. 41, Last annotation update)	
DE	Coagulation factor V precursor (activated protein C cofactor).	
GN	FS, Sus scrofa (Pig).	
OS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Cetartiodactyla; Suidae; Suidae; Sus.	
OK	NCBI_TaxId=9823;	
RN	[1]	
RP	SEQUENCE FROM N.A., AND 3D-STRUCTURE MODELLING OF F5/8 TYPE A AND C	
RP	DOMAINS.	
RP	TISSUE=Liver;	
RA	MEDLINE=21121490; PubMed=11229814;	
RA	Grimm D.R., Colter M.B., Braunschweig M., Alexander L.J., Neame P.J.,	
RA	Kim H.K.W.;	
RT	"Porcine factor V: cDNA cloning, gene mapping, three-dimensional	
RT	protein modeling of membrane binding sites and comparative anatomy of	
RT	domains."	
RU	Cell. Mol. Life Sci. 58:148-159 (2001).	
CC	-1- FUNCTION: Coagulation factor V is a cofactor that participates	
CC	with factor Xa to activate prothrombin to thrombin	
CC	-1- SUBUNIT: Factor Va is composed of a heavy chain and a light	
CC	chain, noncovalently bound. The interaction between the two chains	
CC	is calcium-dependent.	
CC	-1- DOMAIN: Domain B contains 41 x 9 AA tandem repeats. Domains C1	
CC	and C2 may be involved in membrane binding.	
CC	-1- PTM: Thrombin activates factor V proteolytically to the active	
CC	cofactor, factor Va (formation of a heavy chain at the N-	
CC	terminus and a light chain at the C-terminus).	
CC	-1- SIMILARITY: CONTAINS 3 F5/8 TYPE A DOMAINS; EACH IS COMPOSED OF	
CC	2 PLASIOCYANIN-LIKE REPEATS.	
CC	-1- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.	
CC	-1- SIMILARITY: STRONG. TO COAGULATION FACTOR VII.	
CC	-----	
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration	
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -	
CC	the European Bioinformatics Institute. There are no restrictions on its	
CC	use by non-profit institutions as long as its content is in no way	
CC	modified and this statement is not removed, usage by and for commercial	
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/	
CC	or send an email to license@isb-sib.ch).	
CC	-----	
DR	EMBL: AF191308; IACG28381.1; -.	
DR	HSSP: P12259; ICET.	
DR	InterPro: IPR001117; Cu-oxidase.	
DR	InterPro: IPR00421; FA58.C.	
DR	Pfam: PF00394; Cu-oxidase; 3.	
DR	Pfam: PF00754; F5-F8 type-C; 2.	
DR	SMART: SM00231; FA58C; 2.	
DR	PROSITE: PS00079; METALLOPROTEIN_OXIDASE1; 2.	
DR	PROSITE: PS01285; FA58C.1; 2.	
DR	PROSITE: PS01286; FA58C.2; 2.	
KM	Blood coagulation; glycoprotein; Sulfation; Calcium; Signal; Zymogen;	
KM	Repeat.	
FT	SIGNAL	1 22
FT	CHAIN	23 2258
FT	CHAIN	23 737
FT	PEPTIDE	738 1611
FT		
FT	CHAIN	1612 3258
FT	DOMAIN	30 229
FT	DOMAIN	30 193
FT	DOMAIN	203 329
FT	DOMAIN	348 683
FT	DOMAIN	348 525
FT	DOMAIN	535 683
FT		
FT	POTENTIAL.	
FT	COAGULATION FACTOR V.	
FT	HEAVY CHAIN (BY SIMILARITY).	
FT	ACTIVATION PEPTIDE (CONNECTING REGION)	
FT	(BY SIMILARITY).	
FT	LIGHT CHAIN (BY SIMILARITY).	
FT	F5/8 TYPE A 1.	
FT	PLASIOCYANIN-LIKE 1.	
FT	PLASIOCYANIN-LIKE 2.	
FT	F5/8 TYPE A 2.	
FT	PLASIOCYANIN-LIKE 3.	
FT	PLASIOCYANIN-LIKE 4.	

FT	DOMAIN	691	1611	B1 X 9 AA APPROXIMATE TANDEM REPEATS OF
FT	DOMAIN	168	1539	T-L-S-P-D-L-[GS]-[HQ]-T.
FT	REPEAT	1168	1176	1.
FT	REPEAT	1177	1185	2.
FT	REPEAT	1186	1194	3.
FT	REPEAT	1195	1203	4.
FT	REPEAT	1204	1212	5.
FT	REPEAT	1213	1221	6.
FT	REPEAT	1221	1230	7.
FT	REPEAT	1231	1239	8.
FT	REPEAT	1240	1248	9.
FT	REPEAT	1249	1257	10.
FT	REPEAT	1258	1266	11.
FT	REPEAT	1267	1275	12.
FT	REPEAT	1276	1284	13.
FT	REPEAT	1285	1293	14.
FT	REPEAT	1294	1302	15.
FT	REPEAT	1303	1311	16.
FT	REPEAT	1312	1320	17.
FT	REPEAT	1321	1329	18.
FT	REPEAT	1330	1338	19.
FT	REPEAT	1339	1347	20.
FT	REPEAT	1348	1356	21.
FT	REPEAT	1357	1365	22.
FT	REPEAT	1366	1374	23.
FT	REPEAT	1375	1383	24.
FT	REPEAT	1384	1392	25.
FT	REPEAT	1393	1401	26.
FT	REPEAT	1402	1410	27.
FT	REPEAT	1411	1419	28.
FT	REPEAT	1420	1428	29.
FT	REPEAT	1429	1437	30.
FT	REPEAT	1438	1446	31.
FT	REPEAT	1447	1455	32.
FT	REPEAT	1456	1464	33.
FT	REPEAT	1465	1473	34.
FT	REPEAT	1474	1482	35.
FT	REPEAT	1483	1491	36.
FT	REPEAT	1492	1500	37.
FT	REPEAT	1501	1509	38.
FT	REPEAT	1510	1518	39.
FT	REPEAT	1519	1527	40.
FT	REPEAT	1531	1539	41.
FT	DOMAIN	1616	1941	F5/8 TYPE A 3.
FT	DOMAIN	1795	1785	PLASTOCYANIN-LIKE 5.
FT	DOMAIN	1795	1941	PLASTOCYANIN-LIKE 6.
FT	DOMAIN	1942	2095	F5/8 TYPE C 1.
FT	DOMAIN	2100	2255	F5/8 TYPE C 2.
FT	SITE	737	738	CLEAVAGE (BY THROMBIN) (BY SIMILARITY).
FT	SITE	1029	1030	CLEAVAGE (BY THROMBIN) (BY SIMILARITY).
FT	SITE	1611	1612	CLEAVAGE (BY THROMBIN) (BY SIMILARITY).
FT	DISULFID	167	193	PROBABLE.
FT	DISULFID	499	525	PROBABLE.
FT	DISULFID	1759	1785	PROBABLE.
FT	DISULFID	1941	2095	BY SIMILARITY.
FT	DISULFID	2100	2255	BY SIMILARITY.
FT	MOD_RES	692	692	SULETION (POTENTIAL).
FT	MOD_RES	696	696	SULETION (POTENTIAL).
FT	MOD_RES	724	724	SULETION (POTENTIAL).
FT	MOD_RES	726	726	SULETION (POTENTIAL).
FT	MOD_RES	745	745	SULETION (POTENTIAL).
FT	MOD_RES	1560	1560	SULETION (POTENTIAL).
FT	MOD_RES	1576	1576	SULETION (POTENTIAL).
FT	MOD_RES	1581	1581	SULETION (POTENTIAL).
FT	MOD_RES	1584	1584	SULETION (POTENTIAL).
FT	MOD_RES	1588	1588	SULETION (POTENTIAL).
FT	MOD_RES	1631	1631	SULETION (POTENTIAL).
FT	MOD_RES	1631	1631	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBHYD	253	253	
FT	CARBHYD	255	255	

Query Match	Best Local	Similarity	30.4%	Score	2774.5	DB 1:	Length	2258:
Matches	768:	Conservative	367:	Mismatches	875:	Indels	497:	Gaps
FT CARBOHYD	239	239	N-LINKED (GLCNAC. . .)	(POTENTIAL)				
FT CARBOHYD	297	297	N-LINKED (GLCNAC. . .)	(POTENTIAL)				
FT CARBOHYD	382	382	N-LINKED (GLCNAC. . .)	(POTENTIAL)				
FT CARBOHYD	460	460	N-LINKED (GLCNAC. . .)	(POTENTIAL)				
FT CARBOHYD	467	467	N-LINKED (GLCNAC. . .)	(POTENTIAL)				
FT CARBOHYD	553	553	N-LINKED (GLCNAC. . .)	(POTENTIAL)				
FT CARBOHYD	741	741	N-LINKED (GLCNAC. . .)	(POTENTIAL)				
FT CARBOHYD	752	752	N-LINKED (GLCNAC. . .)	(POTENTIAL)				
FT CARBOHYD	760	760	N-LINKED (GLCNAC. . .)	(POTENTIAL)				
FT CARBOHYD	776	776	N-LINKED (GLCNAC. . .)	(POTENTIAL)				
FT CARBOHYD	782	782	N-LINKED (GLCNAC. . .)	(POTENTIAL)				
FT CARBOHYD	899	899	N-LINKED (GLCNAC. . .)	(POTENTIAL)				
FT CARBOHYD	960	960	N-LINKED (GLCNAC. . .)	(POTENTIAL)				
FT CARBOHYD	1048	1048	N-LINKED (GLCNAC. . .)	(POTENTIAL)				
FT CARBOHYD	1057	1057	N-LINKED (GLCNAC. . .)	(POTENTIAL)				
FT CARBOHYD	1066	1066	N-LINKED (GLCNAC. . .)	(POTENTIAL)				
FT CARBOHYD	1073	1073	N-LINKED (GLCNAC. . .)	(POTENTIAL)				
FT CARBOHYD	1089	1089	N-LINKED (GLCNAC. . .)	(POTENTIAL)				
FT CARBOHYD	1174	1174	N-LINKED (GLCNAC. . .)	(POTENTIAL)				
FT CARBOHYD	1480	1480	N-LINKED (GLCNAC. . .)	(POTENTIAL)				
FT CARBOHYD	1537	1537	N-LINKED (GLCNAC. . .)	(POTENTIAL)				
FT CARBOHYD	1597	1597	N-LINKED (GLCNAC. . .)	(POTENTIAL)				
FT CARBOHYD	1737	1737	N-LINKED (GLCNAC. . .)	(POTENTIAL)				
FT CARBOHYD	1866	1866	N-LINKED (GLCNAC. . .)	(POTENTIAL)				
FT CARBOHYD	2044	2044	N-LINKED (GLCNAC. . .)	(POTENTIAL)				
FT CARBOHYD	2243	2243	N-LINKED (GLCNAC. . .)	(POTENTIAL)				
SQ SEQUENCE	2258 AA:	256078 MW:	9159B9E0076A2ACC	CRC64:				
Query Match	Best Local	Similarity	30.4%	Score	2774.5	DB 1:	Length	2258:
Matches	768:	Conservative	367:	Mismatches	875:	Indels	497:	Gaps
OY 22 RRYLYGAVELSNQDMSDGLPEVDPAKRRPVKSPRPNTSVYKYKTLVEFDHLFNIA	81							
DB 32 KQFYAAQSISSKWNHDE-----PHPSSSPFATS--FKIYIRAY--EAFQRE	76							
OY 82 KRPRWMLGLPTQIAEYVDYVITLKNMASHPSLSHAAGVSVMKASGAEDYDQTSORE	141							
DB 77 KPRSMSSGLPLPTLYADVDIKVHFRRKDKPLSIHPGITYKSRFAGASYDPHFVLE	136							
OY 142 KEDDVFVFGSGSHYVQVLYKENGPAASPLPLTLYSHVDLVKRLNSGALVLCREG	201							
DB 137 KMDDAVAGQDEYETVWNISSEDSGPHNDPCLPTIYSENIIDPFNSGLGLPLICKRG	196							
OY 202 SLAKKQTL--HFIILFVFEDEKSHSTKNSLMQORDAASAPMKHNVNYSNR	259							
DB 197 TLEEDGIQKMDKQVLYMFAVDESKSNQSS-----LMIYNGTVNG	240							
OY 260 SLRGLGCHKRSYVNVYVIGMGTTPREVHSIFLEGTPLVYRNHQAOLESPITPLTAQTL	319							
DB 241 TMEDTVCAVTHYSIMHLIGMSSGPPELSFTHPSGVLEQNNHNVASITLVSATSTTAMTV	300							
OY 320 MDLGFLFLFCHSISSHODGMEAYVAVDSCPEEPLQRLKNNEAEYDDILTSEMDVAF	379							
DB 301 SPEGKWPISSLIPHFQAGQAVIDIKNCARKTKRPRK-----LYRQD-----	343							
OY 380 DDDNSPSTQIRSYAKKHPTVWVYIAAEEEDMDYVAPLVAFDDRSYKSOYLNNQFORIG	439							
DB 344 -----RHKIKRMEYFIAAEEYIMVAPILPAMMDCKRYSLSLINFNSQIG	388							
OY 440 RYKKYVREMATVDETETFR--EAIOHESGILGLPLGEGDPLTLIFPNQASRPVNIYH	497							
DB 389 KHKYKVVYKQYODESFTKRLBNPNKKEGILGPIYRAQVRDLKLVFNMAASYSISYH	448							
OY 498 GJT-----DVRPLYSRLRPKGVKHLKDFPLTPELFRKTKVYVYEDQSPRDLRTX	551							
DB 449 GTTSPYEDDVAANSSTSDNNMTIR-----AQOPEYITAKMILSDSPEDNADQCLTR	503							
OY 552 YSSVNMERDIASGLGLPLITCYKESVDQRGQNDMSDKRNVILSVEDENRSKYTLTENTQ	611							
DB 504 YYSNDVITRDIASGLIGLLICKSRSLDRQIORTADIDQKAVFVDEKMSWYIEDNTY	563							
OY 612 RLPLNPACVQLEDEPQASINHSINGVEPSL-QLSVLCHFAVAWTLISIGAQTFSLV	670							

```

Db      564  KFCENPEKVRKDDPKEFYENIMSTINGYVESIPILGFCFDYVQWFFCVRTHDNLITL 623
Oy      671  FFSGYTFKHKMYEDTTLFPFSGEYVFMSEMENPGLMILGCHNSDFNRNMTALTKVSSC 730
Db      624  HFTGHSFYIGKHEDTLTLFPMGSESVYVMDVGMVLTMMNSPNRKKLQKPRDYKC 683
Oy      731  DKMTG--DYEDSYEDISAYLL-----SKNALIEP-----RSPSONS 765
Db      684  IRDDDESTEIYESSSTTLTRKMHDSSENKEENDEYODDLASVGLKRSFRNS 743
Oy      766  RHPSTRKQKFNATVTP--ENDIEKTDWFAHRTMPMKIQNVSSDLMMLKQSPFPGJSL 824
Db      744  LY--QEDDEFNLTLALENNSEFIP-----STD-----RAVDSNSSSP 780
Oy      825  SDLOEAKYETFSDD--PSPGALDSNNS-----LSEKTHRPQLHSGDMVFTF--ESG 873
Db      781  GNISRAPANTFTEPKLILPHEATKAGSPRRHGLVKMLVLRRRHQSDPYSEDPTEMP 840
Oy      874  LQ-----LRLNE-----KLGTATE--LAKIDFVYSTSNMLISTIP 909
Db      841  LQSYTGISLFPFEGEGRNREHPKRRFKAAGDQAAHRSQMEFPAHKTGMH-----IS 896
Oy      910  SDMLAGTDNTSLGPPMPVHYDSOLDTTLFGKSSPLTESGGPISLSEENN----- 962
Db      897  QDN-----SSSSSMGPLE-----DLSSDLLLERK--DSTINGKMHLYSEKGYEIVODA 945
Oy      963  DSKLESGLMANSQES--SMGKNVSTESGRLFGKRAHGPALLTKDNALFKVYSILTK 1019
Db      946  DEDAVNKLPPNPNQASRSGENIPFNK-----HGKORHPIFYVTRH-----LLQGE 993
Oy      1020  NKTSNNS-----ATNRKTHIDGSPILLIENSPVOMNLT--ESDPEFKVYPLHDMN 1069
Db      994  RQDEGNSILKGRFTRIRKKKKRKYVHVPLSPRSFNPRLGKANTPFS----- 1043
Oy      1070  LMDKNTALRLMHMSNKTSSKNEMVQAKKESRPIPPQAN--PDSFPRKALFPEBSR 1126
Db      1044  --DRONHSLHSENER-----FPPDYLKQTFPSMNLISLSDH-- 1082
Oy      1127  WIOTRGKNSLN--SGQGPSRQVLSLGPESVGNQFLSEKNKYVVGKGFYDVGLKEM 1185
Db      1083  ----HNQNLPMDFQTGSSP-----LDLYOT 1103
Oy      1186  VFPS--RNLFLNLDNHNNTNOEKKIOEIEKKEKTLQENYVLPOLHTVGTGKNM 1243
Db      1104  VTDEPYQTAPIODDPTHSTAVPSHSSLPET-----QNH----- 1140
Oy      1244  KNLFLSTRONVGSYDCAVAPYLQDFRSLNDSTNFKKHTAHFS--KGESENLGIG 1300
Db      1141  ----DYD-----LRNKAAPTIVSEMFSLKLKAGHRTSPDL- 1173
Oy      1301  NOTQIYEVYACT-----TRISPTSOQNFYQSKRKALKQFRLPLETEL-----EK 1348
Db      1174  NOTSLPELSQTTLSPPDGHWTLSPDLSQTTLSPDLSHTTSP--DLGHHTTSLPDLSHT 1230
Oy      1349  RIIVDOTSROMSKNKHLPSTLQIDYNEKKGATQSPSLDCITRSHGISIPANSRPLP 1408
Db      1231  TILSPDLSCOTTLSPDLSHTLS-----PDGHHTTSLPDLSHTTSLSPDGHHTTSP-D 1280
Oy      1409  IAKVSFPSIRPIYTRVLFODNSHLLPAASYRKKGSG--VOESSHFLOGAKNNLSLAIL 1467
Db      1281  LSHHTTSLPDGHHTTSLPDGHHTTSLSPDSQTTLSPDGHHTTSLSVSHHTTSLPDLSHTL 1340
Oy      1468  T--LEMTGQREVSLGTSATNSYTYKKVENYVLP-----KPLDPTS-----SKVEL 1513
Db      1341  SPDLSTHTTSLPDGHHTTSLPDLSQTT-----TILSPDLGHHTTSLPDLSHTTSLPDGHHTL 1394
Oy      1514  LPKHAIYKDLPTLETSMNSPGHLDVVEGSLQGTGCAIKKMEANRKPVPFLKATFESS 1573
Db      1395  SP-----DLSHHT--TILSPDLGSHHTLSPD--LQOTTLSLDFEQTTLSPDLSHHTLSSELS 1444
Oy      1574  AKTPSKLL-----DPLAW-----DNHYGTQIKERBMSQEKSPKTAFRK 1613

```

```

Db      1445  HETLSPDLSQVTLSPDLSEIPSPDLMQTTLSSDINETTSLSPDLRQTSPPDPKST-- 1502
Oy      1614  KDTILSLNACESNHAIAINEGONKBEIEVTWAKQGTERTLCSD-----NPPYLRHQ 1666
Db      1503  -----ISESSQSVLLPFGQTSFPDLGDRPSPSHSTLNTFELPREFPDMVYGLS 1554
Oy      1667  RE-----ITRPTLOSQDE-----IDYDTISVEKKKEDDIYDEENOSP----- 1707
Db      1555  RDGQDVLEIPRQGEENSEEDYKIDIVEYDDPYQTDVRT-----DINSNPNPNIAM 1608
Oy      1708  --RSFOKTRHFTFANVERLMDYGASSSHVLRANAQSGSPQ--FKVYVQETPDSGF 1762
Db      1609  YLSSNNGNRNRYLIAELSLMDYSK-----FTQEDIDDPETHYTKVVERKATDSTF 1662
Oy      1763  TOPLYNGELNHTLGLPYIRAEVEDNIWYTERNOASRPSYSLSIYE-----ED 1814
Db      1663  TKLDPRGEYEELHGIILGPIIRAEVDYIVRFRKNLASRPSYSLAHGLSYEKSSEKTYED 1722
Oy      1815  QRGAPRKNPFYKRETYRPMYVQHMAPTKDEPDKAMAFESDQVDEKDYHGLIGPL 1874
Db      1723  DSEPMKEDNAVOPSSITYWHAITERSGPESGACAMAYIVANVPERKHSLGILPL 1782
Oy      1875  LVCHINTLPAHGRQVYOEPALEFTIFDETSYFTENMERKCRAPCNLOMEDPFEKEN 1934
Db      1783  LICRKTGLKRNMFVMDREFVLLFMVFEKKSWMYERKFTRSR-----LTSSEVNS 1836
Oy      1935  YRFHAIYIMDTLPLGYMADORIRMYLLSNGSNENHISHSFGHYVTRKKEKYKAL 1994
Db      1637  HKFHAIYNGIYV-LPGILMYEOEWVRLHLNIGSGSRDILHYVHFHQTLLENGTQOHOGLV 1895
Oy      1995  YNLYPGFETVEMLPSKAGIWRVCLIGELHLAGSTLELYSNKQTPPLGMAAGHIDF 2054
Db      1896  WPLPGSFTELMKTSKAGWMLDTEGENORAGQOTFLIDHECKMPGLSTGLIDS 1955
Oy      2055  QITASQCGOMAPKIAIRHYSGSIKMSKKEPS-----NIVYDLAMTHIGKQSG 2107
Db      1956  QIRASERQWQPKIARLKNNGSINMAYIT--DFTSGESNKKWIVYDQKRYVYTGIGTQSG 2014
Oy      2108  AROKFSLYISOFITMYSLDGKKMQYTRNGSTGLTMEFNGVDSGKIHNIENPILIARY 2167
Db      2015  AKYLLKSYTTEFNVAVSSDORMRKJFKGSRKNMYNGNSDASTTEQDEPPVYARY 2074
Oy      2168  IRLAPHTYSIRSLTLMELMGCDLNSCMPLGMSKRAISDAQITASSTYTMKFT--WSPSK 2226
Db      2075  IRISPESTYNNKPALLELQGEVNGCSTPLGMSGINKNEBITASFSKKSMMGDWEPFR 2134
Oy      2227  ARHLHGSRNARPOYNNPKEMLVDPQKTMKYGVYVQCKSLTLMYNYKKEFLISSSD 2286
Db      2135  ARUNAGQRYNAMAQAKANNMOMQIDILKIKITAITTQGGKSLSSMEYVKRTTYQISDR 2194
Oy      2287  GHOWTLEFQNGKV--KVFQGNQDSFTYVNSLDPPLTRYLRIHPQSVHQAIALMEYLG 2344
Db      2195  GVEKKSYSREKSNMVDKIEGNNNKHGVKNFEPILSRITRIIPKMMQNSIALRLLEFG 2254
Oy      2345  CEAMODLY 2351
Db      2255  C---DIT 2258

```

```

RESULT 6
ID_FAS_BOVIN STANDARD; PRT; 2211 AA.
AC 028107; Q28108;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Coagulation factor V precursor (Activated protein C cofactor).
GN F5.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
NCBI_TaxID=9913;

```

```

RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-Liver;
RA MEDLINE-92147638; PubMed-1737753;
RA Guinlo E.R., Esmon C.T., Mann K.G., Macgillivray R.T.;
RT "The complete cDNA sequence of bovine coagulation factor V.";
RL J. Biol. Chem. 267:2971-2978(1992).
CC -1- FUNCTION: COAGULATION FACTOR V IS A COFACTOR THAT PARTICIPATES
CC WITH FACTOR XA TO ACTIVATE PROTHROMBIN TO THROMBIN.
CC -1- SUBUNIT: Factor Va is composed of a heavy chain and a light
CC chain, noncovalently bound. The interaction between the two chains
CC is calcium-dependent.
CC -1- DOMAIN: Domain B contains 29.5 x 9 AA tandem repeats, and 2 x 14
CC AA repeats.
CC -1- PTM: Thrombin activates factor V proteolytically to the active
CC cofactor, factor Va (formation of a heavy chain at the N-
CC terminus and a light chain at the C-terminus).
CC -1- PTM: SULEFATION IS REQUIRED FOR EFFICIENT THROMBIN CLEAVAGE AND
CC ACTIVATION AND FOR FULL PROCOAGULANT ACTIVITY (BY SIMILARITY).
CC -1- SIMILARITY: CONTAINS 3 F5/8 TYPE A DOMAINS; EACH IS COMPOSED OF
CC 2 PLASTOCYANIN-LIKE REPEATS.
CC -1- SIMILARITY: STRONG, TO COAGULATION FACTOR VIII.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M81440; AAA30512.1; .
DR EMBL; M81441; AAA30513.1; .
DR HSSP; P12259; 1CZT.
DR InterPro; IPR001117; Cu-oxidase.
DR InterPro; IPR000421; FA58_C.
DR Pfam; PF00394; Cu-oxidase; 3.
DR Pfam; PF00754; F5_F8_type_C; 2.
DR SMART; SM00231; FA58C; 2.
DR PROSITE; PS00079; MULTICOPPER_OXIDASE1; 2.
DR PROSITE; PS01285; FA58C_1; 2.
DR PROSITE; PS01286; FA58C_2; 2.
KW Blood coagulation; Glycoprotein; Sulfation; Calcium; Signal; Zymogen;
KW Repeat.
FT SIGNAL; 1 28 POTENTIAL.
FT CHAIN; 29 741 COAGULATION FACTOR V.
FT CHAIN; 29 741 HEAVY CHAIN (BY SIMILARITY).
FT PEPTIDE; 742 1564 ACTIVATION PEPTIDE (CONNECTING REGION)
(BY SIMILARITY).
FT CHAIN; 1565 2211 LIGHT CHAIN (BY SIMILARITY).
FT DOMAIN; 30 327 F5/8 TYPE A 1.
FT DOMAIN; 30 193 PLASTOCYANIN-LIKE 1.
FT DOMAIN; 327 348 PLASTOCYANIN-LIKE 2.
FT DOMAIN; 348 686 F5/8 TYPE A 2.
FT DOMAIN; 525 535 PLASTOCYANIN-LIKE 3.
FT DOMAIN; 535 686 PLASTOCYANIN-LIKE 4.
FT DOMAIN; 686 1564 B.
FT SIMILAR; 899 935 TO 17 AA REPEATS IN HUMAN FAS.
FT DOMAIN; 1124 1151 2 x 14 AA TANDEN REPEATS.
FT REPEAT; 1124 1137 1-1.
FT REPEAT; 1138 1151 1-2.
FT DOMAIN; 1188 1453 30 x 9 AA APPROXIMATE TANDEN REPEATS OF
[As]-L-S-P-D-[Lp]-[Gs]-Q-[Te].
FT REPEAT; 1188 1196 2-1.
FT REPEAT; 1197 1205 2-2.
FT REPEAT; 1206 1214 2-3.
FT REPEAT; 1215 1223 2-4.
FT REPEAT; 1224 1232 2-5.
FT REPEAT; 1233 1241 2-6.
FT REPEAT; 1242 1250 2-7.
FT REPEAT; 1251 1259 2-8.
FT REPEAT; 1260 1268 2-9.

```

```

FT REPEAT 1269 1277 2-10.
FT REPEAT 1278 1286 2-11.
FT REPEAT 1287 1295 2-12.
FT REPEAT 1296 1304 2-13.
FT REPEAT 1305 1313 2-14.
FT REPEAT 1314 1322 2-15.
FT REPEAT 1323 1331 2-16.
FT REPEAT 1332 1340 2-17.
FT REPEAT 1341 1349 2-18.
FT REPEAT 1350 1358 2-19.
FT REPEAT 1359 1367 2-20.
FT REPEAT 1368 1376 2-21.
FT REPEAT 1377 1385 2-22.
FT REPEAT 1386 1394 2-23.
FT REPEAT 1395 1403 2-24.
FT REPEAT 1404 1412 2-25.
FT REPEAT 1413 1421 2-26.
FT REPEAT 1422 1430 2-27.
FT REPEAT 1431 1439 2-28.
FT REPEAT 1440 1444 2-29 (PARTIAL).
FT REPEAT 1445 1453 2-30.
FT DOMAIN 1456 1890 F5/8 TYPE A 3.
FT DOMAIN 1569 1738 PLASTOCYANIN-LIKE 5.
FT DOMAIN 1748 1890 PLASTOCYANIN-LIKE 6.
FT DOMAIN 1894 2048 F5/8 TYPE C 1.
FT DOMAIN 2053 2208 F5/8 TYPE C 2.
FT SITE 742 742 CLEAVAGE (BY THROMBIN) (BY SIMILARITY).
FT SITE 1034 1035 CLEAVAGE (BY THROMBIN) (BY SIMILARITY).
FT SITE 1564 1565 CLEAVAGE (BY THROMBIN) (BY SIMILARITY).
FT DISULFID 167 193 PROBABLE.
FT DISULFID 499 525 PROBABLE.
FT DISULFID 1712 1738 PROBABLE.
FT DISULFID 1894 2048 BY SIMILARITY.
FT DISULFID 2053 2208 SULEFATION (POTENTIAL).
FT MOD_RES 697 697 SULEFATION (POTENTIAL).
FT MOD_RES 701 701 SULEFATION (POTENTIAL).
FT MOD_RES 730 730 SULEFATION (POTENTIAL).
FT MOD_RES 1513 1513 SULEFATION (POTENTIAL).
FT MOD_RES 1529 1529 SULEFATION (POTENTIAL).
FT MOD_RES 1537 1537 SULEFATION (POTENTIAL).
FT MOD_RES 1541 1541 SULEFATION (POTENTIAL).
FT CARBOHYD 225 225 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 235 235 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 297 297 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 382 382 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 460 460 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 553 553 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 587 587 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 745 745 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 756 756 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 774 774 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 780 780 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 902 902 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 952 952 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 964 964 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1044 1044 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1053 1053 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1062 1062 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1071 1071 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1078 1078 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1094 1094 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1451 1451 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1490 1490 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1550 1550 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1690 1690 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1839 1839 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1997 1997 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2196 2196 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARIANT 587 592 NETLA -> T (IN VARIANT 2).
SQ SEQUENCE 2211 AA; 248981 MW; CBBF90B738667C45 CRC64;

```

Query Match 22.1%; Score 2744; DB 1; Length 2211;
 Best Local Similarity 30.2%; Pred. No. 8.6e-135;

Matches 765; Conservative 381; Mismatches 834; Indels 550; Gaps 80;

OY 22 RRYVLAIVELSMQVDDGELPVDARPPVPKSPFPNTSVYKTLFVETDHLNIA 81
DB 32 ROFYVAASIMNNR-----PESTHL-----SKRFETS--FKIYVREI-EAYFOKE 76
OY 82 KRPFPMMGLGPTIOAEVTDYVYITLKMASHAPVSLHVAVSTWKSSEAEYDQTSORE 141
DB 77 KPOSRISGLGPTLYAVGVDMKVAHKKAKPLSIHAQIKYSKFSSEASTDHTLPM 136
OY 142 KEDDVFPVPGSHYVQVYKENGPMASDPLCLTYSLSHVDLYKDLNSGLGALLVREG 201
DB 137 KMDAVAPGQETYEYEWIISHESGPHDPPCLTHIYYSVNLVEDENSGLLGPTLICKG 196
OY 202 SLAKKQTTL--HKFILLFAVEDEKSWHSEFKNSLMQDDAASABAMPKMTVNGYVR 259
DB 197 TLEDEGTQKMEKQHVLMFAVDESKSMNQTS-----LMTVNGYNG 240
OY 260 SLPLGLGCHRSYVYVHIGTTPREHSLFLEGHTFLVRHROASLESPITFLAQTLL 319
DB 241 TMDPTVCAHDHISHLIGMSGPELSIHENGQVLEQNHKISALTLLSATSTANMTV 300
OY 320 MDLQGFLECHISSHQHGMEAVYKVDSCPEEPOLRMKNEEADYDDDLTDEMDVVF 379
DB 301 SPEGRMTIASLIPRHFGQAGVAYIDIKNCAKTRNPK-----LTEDQ----- 343
OY 380 DDNSPSPFIQIRVAKKHPVYVYTAEEEDMVAFLAPDDRGRYKQYLNNGPORIG 439
DB 344 -----RRHIKMEYFIAAEYVIMQVAPIIPANDDKYRSLHLDNFSNIG 388
OY 440 RKYKVRMATDETETFR--EAIQHESGIIGLLGEGVDTLLIFKNOASRPNIYH 497
DB 389 KHKRVYTKQYODDSFKRLDESSGCGELTRIRQVROVPLTIYFKMMASTRSILPH 448
OY 498 GIT-----DVRPLYSRLPKGVKHLKDPILLGGEIRKYKVTYVDEGPKSDPCL 548
DB 449 GYTFSPYDNEVSSSTSGSNTMIRAVR-----PGETYTYKNTLEDETEMDACL 500
OY 549 TRYSSFYNMERDLASGLIPLILICYESVDQNGNOIMSDKRVILFSEYDENRSMYTE 608
DB 501 TRYYSNVDTIRDLASGLILILICKSRSLDRGJORAADIEGOAVFAVDEKSMYTED 560
OY 609 NIOBELPNAGVQLEDEFOASNIH-----SINGYVFDLSQ--LSVCHLEVAYVWILSIG 662
DB 561 NIYKCEBPYKNDKDFYESNIMNFTLPAINGYVESIPILGCFDDIYQVHFCVIG 620
OY 663 AQTDLVSFSGGYPKKRVYEDTLTLFPFSGEYVFMSPENGLMILGCHSDFRKGMT 722
DB 621 TQNDILTHFGHSFTYKGRHEDTLTLFPMQGESVYTMQVGTWMLTMNNSPRSKLR 680
OY 723 ALIKVSSCDKNTGYDEDSYEDI-----SAYILSKNNA----- 755
DB 681 LRFROAKCIRNDD-----DSYEIIYEPGSGTAMTKKIHDSSEIDEDNDADSYODELALI 737
OY 756 IEPSPSONSRHPTROKOFNATIP--ENDIKTPWFHHTPMPKIKONVSSDILMLLR 814
DB 738 LGLRSFRNSL--NOEDELMTALALEKDSFTIP--SANS-----LDSNSS----- 783
OY 815 QSPFPHGLSLDQBAKYTFSDPSPGA-----TDSNNSLEMTHFRPOLHSG----- 864
DB 784 RSHVSRILAKMFASBLTLLHLAPAGSPLEHAGLDKNSALN-----PMAHSSPYSE 838
OY 865 ----DMVFTPEBSGLD-----RLNEKLGTTAATELKLKLDFFVYSTSNLL 904
DB 839 DPREDHPLSDVGVSLLPFGTGFKNNKPAKHQRFQVGRGAQAKKHKSQTRFPAHTRRL 898
OY 905 ISTIPSDNLAAQTQNTS--LGPSPHVIYDQDLTTLFCKKS-----PLRESGP 954
DB 899 -----SODNSSSSKMP-----WEDIPSLILLQOKDYKILNGEMHLYSEKGS 942
OY 955 LSLSEENDSKLLESGLMNSQES--SMGKNVSTGESRLPKGRAHGPALLT----- 1004
DB 943 YEIIODANENKTVNK-LPNSPONDSTFGEINIPFNS-----HGKOSGHPFTFLVTRRKLQ 997

OY 1005 ----KQNALFEKVISILKTKNTKSNNSATNRHTIDGSPLLIENSPPVWQ-----NILES 1054
DB 998 DROBRNRSLKEGLPLIR-----TRRKKKEKPAVHVLPSRSPHPLRGVNAFS 1048
OY 1055 DFEKRYKTPPLIHRMLMDKATALNLSMKTSSKNMEMQOQKKEPIPPDQNPDS 1114
DB 1049 DKRRN-----HSLHLASNETSLID--LMQTPPSMNLSTA-----ASLPD-- 1087
OY 1115 FFKMLFLESANMIOHTGKNSLN--SQGSPKQOLY--SLGPEKSVBQONLSBKRVVY 1171
DB 1088 -----HOQTSPPNJTQTSPPDLYPVYSPREHYQ----- 1117
OY 1172 GKERTKDVGLKEWEPSPSRNLTNLNLTENNTHNOEKIOEIEERKTLIOENVYLP 1231
DB 1118 -----IFP-----IODSDPYHSTYAPSNSPD 1139
OY 1232 QIHTVGTKNFKNLFLSTRONVBSYD--GAVAPYLDPRSL----- 1273
DB 1140 PHTSTYAPSNSP-----PPOSOJPNVDLANRAIPRDVSOIFPSLELVQWQTSLDS 1194
OY 1274 NDSTNRKTKHTAFKSKGEBENLGLQOTQIYEKACTRISPTSGQNFYQSKRA 1333
DB 1195 QPSISPDIGOMALSPPQGESLSPDQO-----TSLSPDQESLSPDQOTA 1242
OY 1334 LKOFRLPLEET--ELEKRIYDVTSTQMSKNMHLN-----STYLOIDYNEKKGATQ 1386
DB 1243 LSPD--PQESLSPDQOTALSPDS-----QESLSPDQOTALSPDQGESLSPDQO 1294
OY 1387 SPLSDCLFR--SHSIPQANRSPDLPAVYSSFPRIPIYLTIRLFDONSHLPAAVYRK 1443
DB 1295 TSLSPDLSQESLSPDQOTALSPD-----QESLSPDQOTALSPD-----PQESLSP 1344
OY 1444 DSGVOESSHFLQAKKNNLSLALILEMTGQKREYVSLGTSAINSTYIKRYENYVLRPD 1503
DB 1345 DLGQTSLSPD--QGESLSPDQOTALSPDQESL-----SPD 1381
OY 1504 LPKTS-----GKVELLPEKHYOKDLFPTETSNQSPGHLDEGSLLOGTEGATKMEAN 1558
DB 1382 LGQTSLSPDQOESLSP-----DLGQOTALSP-----DLSQESLSPDL----- 1418
OY 1559 RPKGVPLRVATESSAPSKLIDPLAMDNIHYGIQIREBWKSOEKSEPKTAFFKKDTI- 1617
DB 1419 --GQPL-----SPDLSLESL-----SPDLSQDLQKQTS 1446
OY 1618 -LSLN-----ACESNHAIALINGQKPEIYVYAKQGRTERLCSQ-----NPPYLKR 1665
DB 1447 PLDLQNTSHSESQSLPRLPERQUTPBNADIGOMPSPPDSTLANTFLPERFNPVAVGL 1506
OY 1666 QRE-----ITRTLASDQEE-----IDYDQTSYEMKKEDEFDYDENQSP----- 1707
DB 1507 SHDDDDYIEIIPROKESSEDEYGEFEVAVANDYQTDLT-----DINSRNPDMIAA 1560
OY 1708 ----RSFOKTRHYIYAVERLMDYGGSSPVLNRRQSGS--VPQ--FKKVFOEFT 1758
DB 1561 KVLRSNTNRKRYIYIAEILSMYVS-----KFOQSDVDYVEDYVYKRYERKYL 1611
OY 1759 DGSFQPLRGELNEHGLIGPTIRAEVEDNIMTTFPNQASRPSYSSLSYE----- 1812
DB 1612 DSGFTKLPDQGEYEHLGILGPTIRAEVDVDYIOVRFKNLSKRYSLAHAGLSYESSSEBK 1671
OY 1813 --EDQROGAERKRVFVNPETKYFKVVOHMAPTKDEFCAMAVFSDVLEKDVHSL 1870
DB 1672 TYEDDSPMFKEEDNAIOPNKTYYVYVHATISGEPNPGSACRAMAYVASVNEKDIHSL 1731
OY 1871 IGPLVCHNTLNPANHROYTVOEALFTFIDETKSMYTFENNMENRACAPNTOMEDPT 1930
DB 1732 IGPLILCKRGTLDEKTMPPADMEFVLFMFVDEKKSMYDKDKPTSMRASS-----E 1785
OY 1931 FKENTRFPAINGYIMDTLPGLVVAQOQRIKRTILSKNSNENHSHFSCHYVYAKKEEY 1990
DB 1786 VNSHFEPAINGMYTN-LPGLRMYEDVEMRHLNLNLSGSDIHYVHFHQTLLENGTQOH 1844

```

QY 1991 KMALYNLCVCPETVEMLPKRAQIMRWECILGHEHLAAGSTLFLVYSNKKQCTPLGASGH 2050
DB 1845 QLGWPLPQSEFTLEMLKRAKRGWMLDTEYGIQRAQKQTEFLVDRCKMPGJSTGL 1904
QY 2051 IRDFQITASQYQGMAPKRLARLYHSGSINM-----STK-EPFSWIKVDLAPMIIHGIR 2104
DB 1905 IADSQIOASEFMQWYMPKRLARLNGGSYNAMVIEKISTEPENPEIOWDMQKEVLLTGIO 1964
QY 2105 TQGARQKFSLLYSQFIYISLDGKKNQYRGNSGTGLVAFPGVNDSSGKHNIFNPIT 2164
DB 1965 TQGAHKLKPYTTEECVAYSILDRNNRIRKGNSTNVMVYFGGNDASSTIREKNDIDPPVY 2024
QY 2165 ARYRLPHRYHSIRSLRHELMGGLNSCMPLGMSKALSDAQTSAYFTNMFAT-WS 2223
DB 2025 ARYIRISPTGSYKPKRLRLEQCEVNGCSTPLGMSGKLEKQITLASSFRKSMGNYME 2084
QY 2224 PSKARHLQGRNAMPPOVNNPKELQVDQKMKYTGTTQGVKSLTSMTYKEFLISS 2283
DB 2085 PFLARLNAQGRVAMQAKANNMNMQLDILKIKITLAVTGCKSLSSSEKYKSTYTHY 2144
QY 2284 SQDGHQWTLFFQNGKV--KVFQGNQDSFTEPVNSLDPPLLFYRLRIRHQSVMQIALRME 2341
DB 2145 SDOGTDMKPYREKSSWVKIEGNNVNRGHVKNFNPITISFRIRIIPKWNOSIALRLE 2204
QY 2342 VLGCEAQDLY 2351
DB 2205 LFGC---DMY 2211

RESULT 7
CERU_HUMAN STANDARD; PRT: 1065 AA.
ID CERU_HUMAN P00450; Q14063;
DT 21-JUL-1986 (Rel. 01, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Ceruloplasmin precursor (EC 1.16.3.1) (Ferroxidase).
GN CP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN 11
RP SEQUENCE FROM N.A.
RX MEDLINE=86253937; PubMed=2873574;
RA Koschinsky M.L., Funk W.D., van Oost B.A., McGillivray R.T.A.;
RT Complete cDNA sequence of human preceruloplasmin.*;
RL Proc. Natl. Acad. Sci. U.S.A. 83:5086-5090(1986).
RN 12
RP SEQUENCE OF 1-1006 FROM N.A.
RX MEDLINE=95217183; PubMed=7702601;
RA Daimon M., Yamatani K., Igarashi M., Fukase N., Kawanami T.,
RA Kato T., Tomihaga M., Sasaki H.;
RT "Fine structure of human ceruloplasmin gene.*";
RL Biochem. Biophys. Res. Commun. 208:1026-1035(1995).
RN 13
RP SEQUENCE OF 1-40; 549-599; 784-829 AND 919-952 FROM N.A.
RX MEDLINE=86275241; PubMed=375405;
RA Mercer J.F.B., Grimes A.;
RT "Isolation of a human ceruloplasmin cDNA clone that includes the N-
terminal leader sequence.*";
RL FEBS Lett. 203:185-190(1986).
RN 14
RP SEQUENCE OF 218-1065 FROM N.A.
RX MEDLINE=86505876; PubMed=3486416;
RA Yang F., Naylor S.L., Lum J.B., Cutshaw S., McCombs J.L.,
RA Nabershaus K.H., McGill J.R., Adrien G.S., Moore C.M., Barnett D.R.,
RA Bowman B.H.;
RT "Characterization, mapping, and expression of the human ceruloplasmin
gene.*";
RL Proc. Natl. Acad. Sci. U.S.A. 83:3257-3261(1986).
RN 15
RP SEQUENCE OF 20-1065.

RX MEDLINE=84119493; PubMed=6582496;
RA Takahashi N., Ortel T.L., Putnam F.W.;
RT "Single-chain structure of human ceruloplasmin: the complete amino
acid sequence of the whole molecule.*";
RL Proc. Natl. Acad. Sci. U.S.A. 81:390-394(1984).
RN 16
RP SEQUENCE OF 158-333; 518-724 AND 858-1065.
RX MEDLINE=83117800; PubMed=6571985;
RA Takahashi N., Bauman R.A., Ortel T.L., Dwulet F.E., Wang C.-C.,
RA Putnam F.W.;
RT "Internal triplication in the structure of human ceruloplasmin.*";
RL Proc. Natl. Acad. Sci. U.S.A. 80:115-119(1983).
RN 17
RP SEQUENCE OF 501-905.
RX MEDLINE=81193407; PubMed=6940148;
RA Dwulet F.E., Putnam F.W.;
RT "Complete amino acid sequence of a 50,000-dalton fragment of human
ceruloplasmin.*";
RL Proc. Natl. Acad. Sci. U.S.A. 78:790-794(1981).
RN 18
RP SEQUENCE OF 907-1065.
RX MEDLINE=80137543; PubMed=6987229;
RA Kingston I.B., Kingston B.L., Putnam F.W.;
RT "Primary structure of a histidine-rich proteolytic fragment of human
ceruloplasmin. I. Amino acid sequence of the cyanogen bromide
peptides.*";
RL J. Biol. Chem. 255:2878-2885(1980).
RN 19
RP SEQUENCE OF 907-1065.
RX MEDLINE=80137544; PubMed=6987230;
RA Kingston I.B., Kingston B.L., Putnam F.W.;
RT "Primary structure of a histidine-rich proteolytic fragment of human
ceruloplasmin. II. Amino acid sequence of the tryptic peptides.*";
RL J. Biol. Chem. 255:2886-2896(1980).
RN 110
RP SEQUENCE OF 1007-1061 FROM N.A.
RX MEDLINE=90285218; PubMed=2355023;
RA Sanford J.A., Horton W.A., Cupples R.L., Banfiacio M.J.,
RA "Human ceruloplasmin, tissue-specific expression of transcripts
produced by alternative splicing.*";
RL J. Biol. Chem. 265:10780-10785(1990).
RN 111
RP X-RAY CRYSTALLOGRAPHY (3.1 ANGSTROMS).
RA Zaitseva I., Zaitsev V., Card G., Moshkov K., Bax B., Ralph A.,
RA Lindley P.;
RT "The X-ray structure of human serum ceruloplasmin at 3.1 A: nature of
the copper centres.*";
RL J. Biol. Inorg. Chem. 1:15-23(1996).
RN 112
RP FUNCTION: CERULOPLASMIN IS A BLUE, COPPER-BINDING (6-7 ATOMS PER
MOLECULE) GLYCOPROTEIN FOUND IN PLASMA. FOUR POSSIBLE FUNCTIONS
ARE FERROXIDASE ACTIVITY, AMINE OXIDASE ACTIVITY, COPPER TRANSPORT
AND HOMOEOSTASIS, AND SUPEROXIDE DISMUTASE ACTIVITY.
RN 113
RP CATALYTIC ACTIVITY: 4 Fe(2+) + 4 H(+) + O(2) -> 4 Fe(3+) + 2 H(2)O.
RN 114
RP COPACITOR. BINDS 6 CO-IONS PER MOLECULE. THIS PROTEIN BELONGS TO
THE MULTICOPPER OXIDASES WHICH CONTAIN THREE DISTINCT CU CENTERS
KNOWN AS TYPE 1 OR BLUE, TYPE 2 OR NORMAL, AND TYPE 3 OR COPIED
BINDING.
RN 115
RP TISSUE SPECIFICITY: SYNTHESIZED IN LIVER AND SECRETED INTO THE
PLASMA.
RN 116
RP DISEASE: CERULOPLASMIN IS DEFICIENT IN WILSON'S DISEASE.
RN 117
RP SIMILARITY: CONTAINS 3 F5/8 TYPE A DOMAINS; EACH IS COMPOSED OF
2 PLASTOCYANIN-LIKE REPEATS.
RN 118
RP This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb.ch/announce/
or send an email to license@isb.ch).
RN 119
RP EMBL, M13699; AAA51976.1; -

```

DR EMBL: D45045; BAA08085.1; -
 DR EMBL: D45044; BAA08084.1; -
 DR EMBL: D45028; BAA08084.1; JOINED.
 DR EMBL: D45029; BAA08084.1; JOINED.
 DR EMBL: D45030; BAA08084.1; JOINED.
 DR EMBL: D45031; BAA08084.1; JOINED.
 DR EMBL: D45032; BAA08084.1; JOINED.
 DR EMBL: D45033; BAA08084.1; JOINED.
 DR EMBL: D45034; BAA08084.1; JOINED.
 DR EMBL: D45035; BAA08084.1; JOINED.
 DR EMBL: D45036; BAA08084.1; JOINED.
 DR EMBL: D45037; BAA08084.1; JOINED.
 DR EMBL: D45038; BAA08084.1; JOINED.
 DR EMBL: D45039; BAA08084.1; JOINED.
 DR EMBL: D45041; BAA08084.1; JOINED.
 DR EMBL: D45042; BAA08084.1; JOINED.
 DR EMBL: D45043; BAA08084.1; JOINED.
 DR EMBL: D45044; BAA08084.1; -
 DR EMBL: D45045; BAA08085.1; -
 DR EMBL: X04135; CAA27752.1; -
 DR EMBL: X04136; CAA27753.1; -
 DR EMBL: X04137; CAA27754.1; -
 DR EMBL: X04138; CAA27755.1; -
 DR EMBL: M13536; AAA51975.1; -
 DR EMBL: J05506; -; NOT_ANNOTATED_CDS.
 DR PIR: A25443; KUHJ.
 DR PIR: A24165; A24165.
 DR PDB: 1KCW; 12-FEB-97.
 DR GLYCOB: P00450; -
 DR SWISS-2DPAGE: P00450; HUMAN.
 DR SILENCE-2DPAGE: P00450; -
 DR Genew: HGNC:2295; CP.
 DR MIM: 117700; -
 DR MIM: 604290; -
 DR InterPro: IPR001117; Cu-oxidase.
 DR InterPro: IPR002355; Multicu_oxidase2.
 DR Pfam: PF00394; Cu-oxidase; 3.
 DR PROSITE: PS00079; MULTICOPPER_OXIDASE1; 3.
 DR PROSITE: PS00080; MULTICOPPER_OXIDASE2; 1.
 DR KX Oxidoreductase; Copper; Metal-binding; glycoprotein; Plasma; Repeat;
 KM Signal; Polymorphism; 3D-structure.
 FT SIGNAL 1 19
 FT CHAIN 20 1065 CERULOPLASMIN.
 FT DOMAIN 20 357 F5/8 TYPE A 1.
 FT DOMAIN 20 300 PLASTOCYANIN-LIKE 1.
 FT DOMAIN 209 357 PLASTOCYANIN-LIKE 2.
 FT DOMAIN 370 718 F5/8 TYPE A 2.
 FT DOMAIN 370 560 PLASTOCYANIN-LIKE 3.
 FT DOMAIN 570 718 PLASTOCYANIN-LIKE 4.
 FT DOMAIN 730 1061 F5/8 TYPE A 3.
 FT DOMAIN 730 900 PLASTOCYANIN-LIKE 5.
 FT DOMAIN 908 1061 PLASTOCYANIN-LIKE 6.
 FT CARBOHYD 1368 138 N-LINKED (GLCNAc. . .).
 FT CARBOHYD 358 358 N-LINKED (GLCNAc. . .).
 FT CARBOHYD 397 397 N-LINKED (GLCNAc. . .).
 FT CARBOHYD 762 762 N-LINKED (GLCNAc. . .).
 FT DISULFID 174 200 PROBABLE.
 FT DISULFID 276 357 PROBABLE.
 FT DISULFID 534 560 PROBABLE.
 FT DISULFID 637 718 PROBABLE.
 FT DISULFID 874 900 PROBABLE.
 FT METAL 120 120 COPPER (TYPE 2) (BY SIMILARITY).
 FT METAL 122 122 COPPER (TYPE 3) (BY SIMILARITY).
 FT METAL 180 180 COPPER (TYPE 3) (BY SIMILARITY).
 FT METAL 182 182 COPPER (TYPE 3) (BY SIMILARITY).
 FT METAL 994 994 COPPER (TYPE 1) (BY SIMILARITY).
 FT METAL 997 997 COPPER (TYPE 2) (BY SIMILARITY).
 FT METAL 999 999 COPPER (TYPE 3) (BY SIMILARITY).
 FT METAL 1039 1039 COPPER (TYPE 3) (BY SIMILARITY).
 FT METAL 1040 1040 COPPER (TYPE 1) (BY SIMILARITY).
 FT METAL 1041 1041 COPPER (TYPE 3) (BY SIMILARITY).
 FT METAL 1045 1045 COPPER (TYPE 1) (BY SIMILARITY).
 FT METAL 1050 1050 COPPER (TYPE 1) (BY SIMILARITY).

FT VARIANT 79 79 T -> G.
 FT VARIANT 449 449 L -> G.
 FT CONFLICT 1060 1060 E -> EGEEP (IN REF. 4).
 FT SEQUENCE 1065 AA; 122205 MM; 2EF21294ED30F58 CAC64;
 Query Match 10 64; Score 1316; DB 1; Length 1065;
 Best Local Similarity 19.24; Pred. No. 6.5e-61;
 Matches 397; Conservative 203; Mismatches 418; Indels 1054; Gaps 29;
 5 LSTCEFLICLFCEFSATRRYYIGAVELSDMYQSDLGE--LPVDAREPPRPVPSFFPT 61
 6 LGFIFLICTSP-AWAKKHHYIGIETWWDY-ASDHCKELISVDENSNITLQNGDPRI 63
 62 SVYKRTLFVEFTDHLFNINAKPPPMGLPTTQAVVDYTVITLKNASHPVSLAVG 121
 64 GLHYKALYLQYDEFRRTTEKPYWLGFLPTLIKATGDKYVHLKMLASRPYPSHG 123
 122 VSTKASGAEYDQTSQAREKEDDYFGSGSHYVQVLEKNGPMASDPLCLTYSYSHV 101
 124 IYVKEHBAIPNTIDFQRADDKYFEBQITMLATEQSFGSGDNCYTRISHI 103
 182 DLVKDLSGLIGALLVCREGSLAKEKTYOTL-HKFTLLFAVDEGKSWSE-----TKNS 224
 184 DAPKDIASGLIGPLILCKDSDLEKEKHIDREFVMSVVDENFSMTLEDNITKYSEP 243
 235 IMODDDAASARAMPKHTVGVNRSPLQLGCHRRKSYVHVIGMGTPEVSHSLEEGHT 294
 244 EKVDKDNEDPQESNRMYSVNGYFGSLPTLLSCAEDNRYKYLFGMGNEVDVAAPFHQA 303
 295 FLVRNHRQASLEISPIELFAQFLMDLGGFLFCHLSSHQHGMEAYVGVKVSCEBPOL 354
 304 LNKRYRTDITNLEPALFEDAYVVAQNBEMTLSCQNLHLKAGLAEFYQEC----- 357
 355 RKNKNEADYDDLLTDSMDVVRFDONSPEFIQISVAKKPKPTVHTIAAEDMDY 414
 358 ---NKSND-----NIGKKVHR-----YVIAAEELIMY 385
 415 APV-----VLAPDRSYKSOYLNNQPORIGKRYKAFVAYTDEF--KTRFAIQ 462
 386 APGAGIDIFKENTLAPGSDS-AVFEGQCTRIGSGYKKLVREYIDASFTRKRGPGE 443
 463 HESGILGILGAVGGDILLIFKNQSRPNYVPHG-----TDVAPLV---SRSLP 511
 444 EHLGILGVYNAEVDITRTFNKATYLSIEPIGVAFKNNEGITYSPNTNPOSRSVP 503
 512 KGKHLKDPILPEIFRYKVTYVEDEPTSDPCRLTRYSSFVMEKOLASGLIGPL 571
 504 PSASH---VAPEETTYEMTVPKKEVGPTNADPVCLAKMYSAVDPDITGLIGPMK 558
 572 ICYKESVDQRGNQMSDKRNVILFVFDENRSWYLTENIOFLPNPAGVOLDPEFQASN 631
 559 ICKKSLHANGKQDVKEFLPPVYFVDENESTLLEDNIRFETAPDQVCKDEDFQSN 618
 632 IMHSINGVYPSIQ-LVQCHAEVAYTILSIGAQNDELVSFEGSYFFKKMYEDTLFLF 690
 619 KHSNANGFYKNGQGLMCKGDSVWYTLFSAGNEDVDHGYTFSGNTLYMRGRRTALF 678
 691 PSEGETVSMENPGLMILCCHNSDFRNRGATALIKYSSCDKNTGDIYEDSDIAYVL 750
 679 PQTSLIHHMPTGEGTNVCLTTHDYTGMMQKTYVQC----- 718
 751 SKNNALPEPSQNSRHPSTROKQFNAVITIPENDIEKTDPFWAFHRTPMKQNVSSDLL 810
 719 ----- 718
 811 MLKROSPFPHGLSLSDLEAKYETESDDSPADISNNLSLMTHTRRPQLHSGMVTTP 870
 719 --RROSE-----DSIFYLGEERTY----- 735
 871 ESGIQLRLNEKLGTTAATELKKIDFKVSVSTSNLISIPSDNLAACTDNTSSIGGPMSPV 930

```

Db 736 -----IAAYVE----- 742
QY 931 HYDSQDPTLFGCKSSPLTBESGPLSLSEBNDKSLBESGLANSOESSKGNVSTESR 990
Db 743 ----- 742
QY 991 LFKGRAHGPALLTJDNMLFVYSISLTKTKTSNNSATNRKTHIDGSPSLIENSPSWON 1050
Db 743 -----ND- 744
QY 1051 ILBESDTEFKKVPYLIDHMLDKKATATRLNHSNKTTSKNNMWOOKKEGPIPPDAON 1110
Db 745 ----- 744
QY 1111 PDMSFFKMLFLPESARWORTGHKNSLNSGSGSPKQVLSLOPEKSEVGQNFLEKKNV 1170
Db 745 ----- 744
QY 1171 VGKGFTKQVGLKENVPSSRNFL/MLDNLHENTHNOEKKIOBEIEKKEILLQENVVL 1230
Db 745 ----- 744
QY 1231 POHTVYGTKNFKMLFLSTRONVEGSDGAYAPVLODFRSLNDSTNRKRTAFHSK 1290
Db 745 -----YSP----- 747
QY 1291 GEEENLEGLNQKQIVKACTTRISPNTSQGNFVTOFSKRAKLPRLPLETELEKRI 1350
Db 748 -----QR----- 749
QY 1351 IYDSTQSKNNKHLTPBTLQIDYNEKKAITOSPISDCLTSHSIPQANRSLPRLA 1410
Db 750 -----EMEKELHHL----- 758
QY 1411 RVSSFPRIPIYTRVLQDNNSHLPAASYRKDKSGVOSSHPLQAKNNLSIALITTE 1470
Db 759 -----OEQVNSNAFL----- 768
QY 1471 MTGDQREVSLGTSATNSVYKKEVNVLPKPDLPKTSQKVELLPKVAHYQKOLPETS 1530
Db 769 ----- 768
QY 1531 NGSPGHLDVBSLLQGTGALKNNENRGPYFLVATESSAKTSPKLLDPLANDNH 1590
Db 769 ----- 768
QY 1591 GTQIPKEEMKSQKSPKTAFRKKDPTILSNACESNHAIAINEGQNKPELEVTAQGR 1650
Db 769 ----- 768
QY 1651 TERLCSQNPVLKRHOREITFTTLQSQOEIDYDFTISEMKKEDFDIYDENQSPRSF 1710
Db 769 -----DKGE----- 772
QY 1711 OKTRHYFIAYVERLDYGMSSPHVLRNKAQSGSVQPKRYVROEFTDGSFTOPLYRGE 1770
Db 773 -----FTGS----- 800
QY 1771 LNEHLGLGPYIRAEVEDINWTFPNOASRPYSFSSLSLYEBDQROGAERKKNVY- 1827
Db 801 EEBHGLIGPOLHADVDGDKYKILFFKNMATRYSIHA-----HCVQSESTVYPTL 850
QY 1828 PNEKTYFMKVYOHMAPTDEDFCKAMAYFSDVDLEKDVHSGILPLVCTNTLNPAHG 1887
Db 851 PGELTYVMKIPERSAGTEGDSACIPRAYYSJVQVMDYSGIOLPLVCRPRPLAKYFN 910
QY 1888 ROYVVOEFALEFIIPEKTSWYETENMEMRNRBAPCNOMEDPTFRENRPAINGYNDT 1947
Db 911 RKKL--EFALLFLVPEBNESSWILDNIKTYSDREKYNKDDEFTESKMAHMGREFGN 968
QY 1948 LFGVMAQDQKIRKYLISGSNENHSHSGHVFYVYRKKEBKALYNLYPGVFETVM 2007
Db 969 LQGLTMHVGDEVNMYLMGNGNEIDLTVHFGHSFOKRGVYSSDVEDIFFPGYGTILEM 1028

```

```

QY 2008 LPSKAGINRVECLIEGHVHAGNSTFLYYSNK 2039
Db 1029 PFRPGIWLHCHVTDHIMGMETTYVQNE 1060

RESULT 8
CERU_RAT STANDARD; PRT; 1059 AA.
AC P13635; 064719;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Ceruloplasmin precursor (EC 1.16.3.1) (Ferroxidase).
GN CN
OS Rattus norvegicus (Rat)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Sprague-Dawley; TISSUE=Lung, and Liver;
RA MEDLINE-90237081; PubMed-2332446;
RX Fleming R.E., Gilpin J.D.;
RT "Primary structure of rat ceruloplasmin and analysis of
RL tissue-specific gene expression during development.";
RN J. Biol. Chem. 265:7701-7707(1990).
RN [2]
RP SEQUENCE OF 257-294; 571-612 AND 823-892 FROM N.A.
RC TISSUE=Liver;
RX MEDLINE-87137545; PubMed-3818625;
RA Aldred A.R., Grimes A., Schreiber G., Mercer J.F.B.;
RT "Rat ceruloplasmin. Molecular cloning and gene expression in liver,
RL choroid plexus, yolk sac, placenta, and testis.";
RX J. Biol. Chem. 262:2875-2878(1987).
CC -1- FUNCTION: CERULOPLASMIN IS A BLUE, COPPER-BINDING (6-7 ATOMS PER
CC MOLECULE) GLYCOPROTEIN FOUND IN PLASMA. FOUR POSSIBLE FUNCTIONS
CC ARE FERROXIDASE ACTIVITY, AMINE OXIDASE ACTIVITY, COPPER TRANSPORT
CC AND HOMEOSTASIS, AND SUPEROXIDE DISMUTASE ACTIVITY.
CC -1- FUNCTION: MAY ALSO PLAY A ROLE IN FETAL LUNG DEVELOPMENT OR
CC PULMONARY ANTIOXIDANT DEFENSE.
CC -1- CATALYTIC ACTIVITY: 4 Fe(2+) + 4 H(+) + O(2) = 4 Fe(3+) + 2 H(2)O.
CC -1- COFACTOR: BINDS 6 CU-IONS PER MOLECULE. THIS PROTEIN BELONGS TO
CC THE MULTICOPPER OXIDASES WHICH CONTAIN THREE DISTINCT CU CENTERS
CC KNOWN AS TYPE 1 OR BLUE, TYPE 2 OR NORMAL, AND TYPE 3 OR COUPLED
CC BINUCLEAR.
CC -1- TISSUE SPECIFICITY: SYNTHESIZED IN LIVER AND SECRETED INTO THE
CC PLASMA. ALSO CHOROID PLEXUS, YOLK SAC, PLACENTA, AND TESTIS; NOT
CC IN STOMACH AND SMALL INTESTINE. FETAL LUNG AND LIVER.
CC -1- INDUCTION: BY INFLAMMATION.
CC -1- SIMILARITY: CONTAINS 3 F5/8 TYPE A DOMAINS; EACH IS COMPOSED OF
CC 2 PLASTOCYANIN-LIKE REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; L33869; AAA40917.1; -
DR EMBL; M80529; AAB65820.1; -
DR EMBL; J02670; AAA40914.1; ALT_SEQ.
DR EMBL; M14102; AAA40915.1; -
DR PIR; A35210; A35210.
DR PIR; A29564; A29564.
DR HSSP; P00450; 1KCW.
DR INTERPRO; IPR001117; Cu-oxidase.
DR INTERPRO; IPR002355; MulticCu-oxidase2.
DR Pfam; PF00396; Cu-oxidase; 3.
DR PROSITE; PS00079; MULTICOPPER_OXIDASE1; 3.
DR PROSITE; PS00080; MULTICOPPER_OXIDASE2; 1.

```


Db 753 -----QENVSNAFL----- 762
QY 1477 EVGSLGTSATNSVTKKENTVLPKPOLPKTSKVELPKPHIYQKDLFTETSNQSPQH 1536
Db 763 ----- 762
QY 1537 LDVLESSLQGTGEGAIKKNNEANRPGKVPPLRVATESAKTPSKLIDPLAMDNHYGTQIPK 1596
Db 763 ----- 762
QY 1597 EEMSKOESPEKTAFFKRDITLSLACENSHAIAMINQGNKPELEVYMAKQGRTERLCS 1656
Db 763 ----- 762
QY 1657 QNPPVLRKHOREITTTLSQDSQEIDYDQETISVEAKKEDFDIYDEDENQSPSFQKTRH 1716
Db 763 -----DKEE----- 766
QY 1717 YETIAVERLMDYGMSSSPHVLNRNAGSGVPOFKKVVFOETDGSFMOPLYRGELNEHLG 1776
Db 767 FFIGS-----KKKKVYREFTDSTFREQVRRAREEHLG 800
QY 1777 LLAGPYIRAEVEDNIMVTERNOASRPYSFSSLSIYSEDOQOGAARPKRFVK---PNETKT 1833
Db 801 MEGPLIHADVGAQKVKYVFKNMATRPYSIRH-----HGVTKSSVVAPTLPGEVPT 850
QY 1834 YFMKVOHMAPTKDEPDCAKAWYFSDVDLEKDVHSGLSIGPLIYC---HTMTLPNARGRY 1890
Db 851 YIMQIPERSAGTEDSPCLIPWATSYVDKXKOLISOLIPLVCRKSYVAVNPK----- 905
QY 1891 TVQEPALPFIIDETKSNVFTENNERNCRAPNCIOMEDPTREKRYRHAIINGTYMDPLG 1950
Db 906 KKEEFLPLVDESNWYLDQNTINTYDPHEKNDNDNEETESKNHAIANGMFCENLOG 965
QY 1951 LVNAQDQIRKMYLLSKGSENIHSHPSGHVFTVRKEEYKALYNTLPYGFVETEMLS 2010
Db 966 LTMHVSGDEVWYVMAMGEIDLTHTVHFHGHSHFYKRGHISDVDFPQTYOTLEMFPQ 1025
QY 2011 KAGIMRVCELGEBHLHAGKSTLFLVYSNK 2039
Db 1026 TPGTWLHCHVTDHIHAGWTTTYVLPNQ 1054

RESULT 9
CERU_MOUSE STANDARD: PRT: 1062 AA.
ID CERU_MOUSE
AC 061147.
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Ceruloplasmin precursor (EC 1.16.3.1) (Ferroxidase).
GN CP.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Krimp L.W.J., Farhangrazi Z.S., Choi D.W., Gitlin J.D.;
RU Submitted (Mar-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP TISSUE SPECIFICITY.
RX MEDLINE=96294736; PubMed=8690795;
RA Krimp L.W.J., Farhangrazi Z.S., Dugan L.L., Gitlin J.D.;
RT *Ceruloplasmin gene expression in the murine central nervous system.*;
RL J. Clin. Invest. 98:207-215(1996).
CC -1- FUNCTION: CERULOPLASMIN IS A BLUE, COPPER-BINDING (6-7 ATOMS PER
MOLECULE) GLYCOPROTEIN FOUND IN PLASMA. FOUR POSSIBLE FUNCTIONS
CC ARE FERROXIDASE ACTIVITY, AMINE OXIDASE ACTIVITY, COPPER TRANSPORT
CC AND HOMEOSTASIS, AND SUPEROXIDE DISMUTASE ACTIVITY.
CC -1- CATALYTIC ACTIVITY: 4 Fe(2+) + 4 H(+) + O(2) + 4 Fe(3+) + 2 H(2)O.
CC -1- COFACTOR: BINDS 6 CU-IONS PER MOLECULE. THIS PROTEIN BELONGS TO
CC THE MULTICOPPER OXIDASES WHICH CONTAIN THREE DISTINCT CU CENTERS

CC KNOWN AS TYPE 1 OR BLUE, TYPE 2 OR NORMAL, AND TYPE 3 OR COUPLED
CC BRINCLER.
CC -1- TISSUE SPECIFICITY: MANY TISSUES, INCLUDING LIVER, EYE AND BRAIN.
CC -1- SIMILARITY: CONTAINS 3 F5/8 TYPE A DOMAINS; EACH IS COMPOSED OF
CC 2 PLASTOCYANIN-LIKE REPEATS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sb.ch/announce/>
CC or send an email to license@isb-sb.ch).
CC
CC EMBL: U49430; AAB07996.1; -.
CC HSSP: P00430; ICKM.
DR MSO: MGI:88476; CP.
DR InterPro: IPR001117; Cu-oxidase.
DR InterPro: IPR002355; Multicu_oxidase2.
DR Pfam: PF00394; Cu-oxidase; 3.
DR PROSITE: PS00079; MULTICOPPER_OXIDASE1; 3.
DR PROSITE: PS00080; MULTICOPPER_OXIDASE2; 1.
KW Oxidoreductase; Copper; Metal-binding; glycoprotein; Plasma; Repeat;
KW Signal.
FT SIGNAL 1 19 BY SIMILARITY.
FT CHAIN 20 1062 CERULOPLASMIN.
FT DOMAIN 20 356 F5/8 TYPE A 1.
FT DOMAIN 20 199 PLASTOCYANIN-LIKE 1.
FT DOMAIN 208 356 PLASTOCYANIN-LIKE 2.
FT DOMAIN 369 713 F5/8 TYPE A 2.
FT DOMAIN 369 713 PLASTOCYANIN-LIKE 3.
FT DOMAIN 565 713 PLASTOCYANIN-LIKE 4.
FT DOMAIN 725 1057 F5/8 TYPE A 3.
FT DOMAIN 725 896 PLASTOCYANIN-LIKE 5.
FT DOMAIN 904 1057 PLASTOCYANIN-LIKE 6.
FT DISULFID 173 199 BY SIMILARITY.
FT DISULFID 275 356 BY SIMILARITY.
FT DISULFID 529 555 BY SIMILARITY.
FT DISULFID 632 713 BY SIMILARITY.
FT DISULFID 870 896 BY SIMILARITY.
FT METAL 120 120 COPPER (TYPE 2) (BY SIMILARITY).
FT METAL 122 122 COPPER (TYPE 3) (BY SIMILARITY).
FT METAL 179 179 COPPER (TYPE 3) (BY SIMILARITY).
FT METAL 181 181 COPPER (TYPE 3) (BY SIMILARITY).
FT METAL 930 990 COPPER (TYPE 1) (BY SIMILARITY).
FT METAL 930 993 COPPER (TYPE 2) (BY SIMILARITY).
FT METAL 995 995 COPPER (TYPE 3) (BY SIMILARITY).
FT METAL 1035 1035 COPPER (TYPE 3) (BY SIMILARITY).
FT METAL 1036 1036 COPPER (TYPE 1) (BY SIMILARITY).
FT METAL 1037 1037 COPPER (TYPE 3) (BY SIMILARITY).
FT METAL 1041 1041 COPPER (TYPE 1) (BY SIMILARITY).
FT METAL 1046 1046 COPPER (TYPE 1) (BY SIMILARITY).
FT CARBOHYD 138 138 N-LINKED (GLCNAc. . .) (POTENTIAL).
FT CARBOHYD 226 226 N-LINKED (GLCNAc. . .) (POTENTIAL).
FT CARBOHYD 583 583 N-LINKED (GLCNAc. . .) (POTENTIAL).
FT CARBOHYD 625 625 N-LINKED (GLCNAc. . .) (POTENTIAL).
FT CARBOHYD 757 757 N-LINKED (GLCNAc. . .) (POTENTIAL).
FT CARBOHYD 922 922 N-LINKED (GLCNAc. . .) (POTENTIAL).
SQ SEQUENCE 1062 AA; F3F5ED09A338F16 CRC64.
Query Match 10.0%; Score 1236; DB 1; Length 1062;
Best Local Similarity 18.9%; Pred. No. 9, 5e-57;
Matches 392; Conservative 196; Mismatches 433; Indels 1054; Gaps 30;
QY 10 FLCLIRFC-----SATRRYYLGAIVLSWPMYO--SDLGELPVDARPPVPRSPFEN 60
Db 3 FILSLFIFLYSSALANDKRIYFGLTEAVWVYASGTEKRLISVITDEGSRNYLQNGDR 62
QY 61 TSVYVKITLFEVFTLHFNIAKRPMPAGLPTQAVYDVYVITLKNASHPVSLAV 120
Db 63 IGRYKALFFETDGTSTKIDPAMLGSLDIPYKAEVEDKTYVHLKNASRIITFHAAH 122
QY 121 GVSVMASGAEYDDQTSQREKEDRVFGSGHTYVMQVLKENGPAADPLCLITVSYLSH 180

```

Db      123 GVTYKREGEAVYEDNTTDEFORADKVLPGQOYYVLAH-NBSPQEGDSNCTVRIYHSH 181
Oy      181 VDLVDLNSGLIGLALVCRGSLAKEKOTUL-HEFILLFVPEEGSMSEKSNIM- 236
Db      182 VDPADPDASGLIGPLICKKGSLLKEKENIDOEVLMEFSVDENLSMTLENTKTCOSE 241
Oy      237 ---ODRDAASARAMPKMHVNGVYNSRLPLGLGCHRSVYMWVIGMTTPEVHSIFLEGH 293
Db      242 PERVDKDNEDFOESNRYTSINGTYGSLPGLSMAADRVKMTYLGNGNEVDVHSAFFHQ 301
Oy      294 TELVNRHQASIEISPTFLTAOTLLMDLGOFLFCHISSHODGMEAVKVCSPERPQ 353
Db      302 ALTSRNYOTDIINFPAFLIDAVMAQNGVWMLSCONLNLKAGLOAFQVOD- 356
Oy      354 LMKKNEAEYDDDLDSMDVYFRDDNSPSFQIRSVAKKHKPTVHYTAAEEDMD 413
Db      357 ---MKPSKD-----NIRKHYRH-----YITAEVYIMN 383
Oy      414 YAP--LVIAPDOR---SYKSOYLNGBPQIRGKRYKVFMAVDEBT--KTRFALQH 463
Db      384 YAPSGIDFTEKEKTLASGSDSGVFEGATRGSGYKKAAYREYTDGSPFNKREKGPDEE 443
Oy      464 ESGILGPLYGEVDTLLIIFKNQASRPYNIYPHGT---DVRPLYSRRLPKGYKHLMD 519
Db      444 HGLIGLPVYMAEVGDTIKTFPHKQGHLSLOPMGVSTFAMENGFTYGG--PGASSQQA 500
Oy      520 FP--ILPGEIFYKWTYVEDEGPTKSPDRCLTRYSSFYNNERDLASGLIGPLIYKFSV 578
Db      501 ASHVAPKKTFTYEMTVKEMGPTYADPCLSMYSAVDPTKLIIFGLIGPKICKKSSL 560
Oy      579 DQRGNOIMSKRNYILFSVDENRSWYLTENQRLPRPAGVQLEPEREQASINWHSNG 638
Db      561 LADGRQKVDKEEYLFPTVDENESLLDDNIMETHAPQDVYKDEEDFOESKNHSHNG 620
Oy      639 YVFDLSOL-SVCLHNAVWYILSIGAOTDPLSVFSGYTKHKMYEDTLLEPESGTV 697
Db      621 FMYGNOSPHMCLGESIYWTYLSAGNEADVHGYPSGWTYLCGGEERDANLFPKHSJTL 680
Oy      698 FMSMENPGMLTIGCHNSDFRNRMGTALLKVSQCDNTGDEYEDSIDISAYLSKNAIE 757
Db      681 LAMPQKGFVDECLTLDHTYGMKQKVTYNNQCG- -QFDEFTVYL- 724
Oy      758 PRFSQNSRHPSTROKQFNATITPENDIEKTPWHAHTPMPKIQANVSSDMLMLRSP 817
Db      725 ----- 724
Oy      818 TPRGLSLDLOBAKETFPDPSPGAIDSNNLSLSEMTFHRPQLHSGDMVFTPESGLQLR 877
Db      725 ----- 724
Oy      878 LNEKLGTAATELKLKLDKFVSTSNMLISTIPSDNLAGTDNTSLGPPSMVHYSDOLD 937
Db      725 ----- 724
Oy      938 TTLFGKKSPLTESGGLSLEENNDKLLBESGLMNSQESWGKSNVSTESGRLKGRRA 997
Db      725 -----GERT 728
Oy      998 HGPALLTKDNALFKVYSISLTKTKTSNNSATNRKTHIDGPSLIENSPSYMONILSDTE 1057
Db      729 Y-----YVDA----- 733
Oy      1058 FKKVPLIHDRMLDKNATALLRLNHSNKTSSKNMEVYQKKGGPFPDAQNPDMSEFFK 1117
Db      734 ----- 733
Oy      1118 MLFLPESAMWIORTHGKNSLNSGOGSPQOLVSLCEPKSVEGQNFLEKKRYVVGKEFT 1177
Db      734 ----- 733
Oy      1178 KDVLKEMVFPSSRMFLVTLNLDNLHENWTHNOEKKIOEIEKKETLLIQENVVLPOIHTVT 1237

```

```

Db      734 ----- 733
Oy      1238 GTKNFMKNLFLSTRQNVGSGYDGAVALDOPFRSLNDSTNRTKHTAHFSKGEENLE 1297
Db      734 -----VEVEND--YSP----- 742
Oy      1298 GLGNQTKQIYEKACTTTRISPTSQONTVYQSRKALQKQRLPLEETELERITVDTST 1357
Db      743 -----SR 744
Oy      1358 QWSKMKHLPSTLTQIDYNEKEKGAITQPSLDCILTRSHSIPQANRSPLEIAKVSFP 1417
Db      745 AMEKELHBL----- 753
Oy      1418 IPIYLFVLEFDNSSHLPDASARRKDSGVQESSHFLQAKKNLSLAITLLEMTGDOR 1477
Db      754 -----QEQNVS----- 759
Oy      1478 VGSLOTSATNSVYKKVENTVLPKPDLPKTSKVELLPVHYLYQKDLPEFTSNGSPGHL 1537
Db      760 ----- 759
Oy      1538 DLVESSLQGTGEGAIKMNANRPGKVPFLRVATBSAKTIPSKLLPLAMDWHGTQLPKE 1597
Db      760 ----- 759
Oy      1598 EWSQEKSPERTAKFKKDTLLSLNCSNHAIAINEGONKPEIETWAKOGRTERLCSO 1657
Db      760 ----- 759
Oy      1658 NPVLAKRQREITPTTLQSDQDEIDYDITSEVEMKKEDPIDVEDENOSPRSFOKTRHY 1717
Db      760 -----NVLDRKEFTI----- 770
Oy      1718 FTAAYERLMDGSSPHVLRNDAQSGVPOFKKVVQFETDGSFQPL-YRGLNHLG 1776
Db      771 -----GS--KYKKVYRQPTDSSFEQVKRAEDHILG 802
Oy      1777 LIGPYRAVEDNIMYTFNRQASRPYSFSSLSYEBDOQGAEPKKNFV--KPNETKT 1833
Db      803 ILGPIIHANVGDKVYFKKMATRPSIHA-----HGVTSSVYVPLTGEVAT 852
Oy      1834 YPMQOHMAPTKDEPCKMAAFVSQVLDVHSGLIGPLVC--HTNLTNPAHGRQV 1890
Db      853 YTMQIPERSGREDSDACIPMATYSYVRVVDLSGLIGPLICRSYKVPSPK----- 907
Oy      1891 TYOERALEFTTFDETKSWYTEMENRCRACNQMDPTFKENTYFHAINGYIMOTLP 1950
Db      908 KKEFFLLFVFEDESNWYIDNITIKYSEHPEKYNKDNBEELSNMHAINGKMEGMLQ 967
Oy      1951 LYMAODORIRWYLLSMGSNENIHSIFSGHVFYVRKKEEYKALYNLYPGVEYEMLP 2010
Db      968 LTMHVKDEVNMYLMGMKNEIDLTHTVYFHGHSFOYKHNAGVYSSDVFLPGYVQDLEMPQ 1027
Oy      2011 KAGIRVECLIGELHAGMSTLFLVYNNKCPPLG 2045
Db      1028 TPGTWLHCHTVDVHAGMATYTYVLPDQETKSG 1062

RESULT 10
MEGM_RAT
AC P70490;
DT 01-NOV-1997 (rel. 35, Created)
DT 01-NOV-1997 (rel. 35, Last sequence update)
DT 30-MAY-2000 (rel. 39, Last annotation update)
DE lactadherin precursor (Milk fat globule-BSG factor 8) (MFG-BS) (O-
  acetyl Gb3 ganglioside synthase) (AGS) (MFGM).
GN MFG8 OR AGS.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
NCBI_Taxid=10116;

```

```

RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN;
RX MEDLINE-96374422; PubMed-8780713;
RA Ogura K., Nara K., Matanabe Y., Kohno K., Tai T., Sanae Y.;
RT "Cloning and expression of cDNA for O-acetylation of GD3
   ganglioside."
RL Biochem. Biophys. Res. Commun. 225:932-938(1996).
CC -1- FUNCTION: MAY BE INVOLVED IN PHOSPHOLIPID BINDING. SEEMS TO
CC PARTICIPATE IN THE O-ACETYLATION OF GD3 GANGLIOSIDE SIALIC ACID.
CC -1- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY: SPLEEN, LUNG, HEART, BRAIN AND MUSCLE.
CC -1- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D84068; BAA12210.1; -.
DR HSSP: P00740; IEDM.
DR InterPro: IPR000561; EGF-1-like.
DR InterPro: IPR001438; EGF_II.
DR InterPro: IPR000421; FA58_C.
DR Pfam: PF00008; EGF_2.
DR Pfam: PF00754; F5_F8_Type_C_2.
DR PRINTS: PR00010; EGFBL00D.
DR SMART: SM00181; EGF; 2.
DR SMART: SM00231; FA58C_2.
DR PROSITE: PS00022; EGF_1; 2.
DR PROSITE: PS01186; EGF_2; 2.
DR PROSITE: PS01285; FA58C_1; 2.
DR PROSITE: PS01286; FA58C_2; 2.
DR Signal: Glycoprotein; Repeat; EGF-1-like domain; Milk.
FT SIGNAL 1 22
FT CHAIN 23 427
FT DOMAIN 24 61 EGF-LIKE 1.
FT DOMAIN 111 267 EGF-LIKE 2.
FT DOMAIN 272 427 F5/8 TYPE C 1.
FT DISULFID 28 39 F5/8 TYPE C 2.
FT DISULFID 33 49 BY SIMILARITY.
FT DISULFID 51 60 BY SIMILARITY.
FT DISULFID 68 79 BY SIMILARITY.
FT DISULFID 93 96 BY SIMILARITY.
FT DISULFID 73 96 BY SIMILARITY.
FT DISULFID 98 107 BY SIMILARITY.
FT DISULFID 111 267 BY SIMILARITY.
FT DISULFID 254 258 BY SIMILARITY.
FT DISULFID 272 427 BY SIMILARITY.
FT SITE 87 89 CELL ATTACHMENT SITE (POTENTIAL).
FT CARBOHYD 61 61 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 230 230 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 280 280 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 390 390 N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 427 AA; 47413 MW; EA8C8631F3EE6047 CRC64;
Query Match 5.3%; Score 663; DB 1; Length 427;
Best Local Similarity 40.3%; Pred. No. 1,6e-27;
Matches 142; Conservative 61; Mismatches 123; Indels 26; Gaps 6;
QY 2018 ECLIGETLHAGNSTLEFLVSNK-----CQPLMASGSHRDPQIRASQY-- 2062
DB 78 KCLVTEDPORG--DIFETXICQCVGSGIHCELFGLGEGALADQISASSVYMG 135
QY 2063 -----GQAPRLARLHYSSINAW--SRKEPFSIKYDILAPYIIGIKTGCAQKRSILX 2116
DB 136 FMSLQRMGPETALRYITGVNHWYRASSYDKEPFIQVDFLRKRRYSGVATQASAGRAET 195
QY 2117 ISQFTIMSLQCKAKKQOTYRNGSTGLMVFEGANDSSGIIHNIFNPILIARYINLRPHYS 2176

```

```

DB 196 LKTFKVAISLDGKRREFFLODESSTGDKREPMGNODNSIKTKINMFPLEKQYRLYPSVCH 255
QY 2177 INSTLMELMGCDLNSCMPLGASRAISDAQITASSYFT--NMFA-TYSPSKARLHOG 2233
DB 256 RGCTLRFELLCCEJLHCGSEPLGKNTTIDPSQTSASSSYTWMLRFAGVYPIALGDLNOC 315
QY 2234 RSNAMPQVNNPKREMLQVDFQKMKRTGVWVGVSLLTSYVKEFLISSODGHQWTLF 2293
DB 316 KINAWAQSNAKEMQLQVDFGTRKTKTITIGGARDGHHQIYVASYKVAHSDQGVQWTVY 375
QY 2294 FQNGKRVQGNQDSFPVYNSIDPRLRRYRIRHPSWYHOIALMRYGC 2345
DB 376 EDGTSKVPQGLNDNSHKNTYRFPFMAYRYVLPJLSMWRITLLELGC 427
RESULT 11
MFGM_MOUSE STANDARD; PRT; 463 AA.
AC P21956; P97800;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Lactadherin precursor (Milk fat globule-EGF factor 8) (MFG-E8) (MFGM)
DE (Sperm surface protein SP47) (MP47).
GN MFG8
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 23-35.
RC TISSUE-Mammary gland;
RX MEDLINE-91046008; PubMed-2122462;
RA Stubbs J.D., Lekutis C., Singer K.L., But A., Yuzukl D.,
RA Srinivasan U., Parry G.;
RT "cDNA cloning of a mouse mammary epithelial cell surface protein
RT reveals the existence of epidermal growth factor-like domains linked
RT to factor VIII-like sequences."
RT Proc. Natl. Acad. Sci. U.S.A. 87:8417-8421(1990).
RN [2]
RP SEQUENCE OF 23-463 FROM N.A.
RC TISSUE-Testis;
RA Ensslin N.A.;
RL Submitted (NOV-1997) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: MAY BE INVOLVED IN PHOSPHOLIPID BINDING. ZONA PELLUCIDA-
CC BINDING PROTEIN.
CC -1- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY: MAMMARY EPITHELIAL CELL SURFACES AND
CC SPERMATOZOAN.
CC -1- DEVELOPMENTAL STAGE: RNA EXPRESSION IS DETECTABLE IN MAMMARY
CC TISSUE FROM NONPREGNANT ANIMALS & MAXIMAL IN THE LACTATING GLAND.
CC -1- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M38337; AAA39534.1; -.
DR EMBL: Y11684; CAA72380.1; -.
DR PIR: A36479; A36479.
DR HSSP: P00740; IEDM.
DR MGD: MGI:102768; Mfg8.
DR InterPro: IPR000561; EGF-1-like.
DR InterPro: IPR001438; EGF_II.
DR InterPro: IPR000421; FA58_C.
DR Pfam: PF00008; EGF_2.
DR Pfam: PF00754; F5_F8_Type_C_2.

```

Query Match 5.38; Score 657; DB 1; Length 463;
Best Local Similarity 42.28; Pred. No. 3.8e-27;
Matches 135; Conservative 63; Mismatches 110; Indels 12; Gaps 5

OY		2037	SUKCOTPLGMAASHINDPOLTASGY-----GQAPKLARLHHYSGINAM--STKEPS	2088
Dd		145	AARCTGTGGEGGLADSDGISASAYMGFMGLAQWGBELARLYRTGVINAWHASNDRP	204
OY		2089	MKVVDLLAPMLIHGIKTCAGARQFESSLISQIFILMTSLDGKMWQTVRGNSTGLTAVFEGN	2146
Dd		205	WUOVNLTKRMVSGVMTOGRSAGRABEYLKTFKKVAJSLDGRPFETIDJESGD-KEFLGN	263
OY		2149	VDSOSIKININPPILARTIRLRPHHYIBSTLAMELMSCDLNCSMPLGMESHALSDAO	2200
Dd		264	LNNLSLKNNMPLENOYRIURLYPASCHRGTTLEFFELLCGLHLGLEPLFGILKNNTLPDSO	323
OY		2209	ITASSFYR--NMFA-TWSPSKARLHLQGRSNAMPPOYNPKEMLOVDPOKIMKYTGVTQ	2265
Dd		324	MSASSYKTWNLRAFGWYPHGLRLDNOKINAMTAQNSAKEMLOYDGTQRQVYGITQ	383
OY		2266	GYSKLTLTMKYKEELISSQDGHQWTLTEFOGKRKYTPQGQDOSFPVNVNSLDPLZTRYL	2322
Dd		384	GARDGEHLQYESYKYAHSDGVOMTYEEQSSKVFQGNLNDNSHKRNIEKDFMARY	443
OY		2326	RHPQSWNYHOTALREMYLGC	2345
Dd		444	RVLPMYSMNRRITLRELLEG	463
RESULT 12				
ID	MEQM_PIG			
AC	P79385:	STANDARD:	PRT:	409 AA.
DT	01-NOV-1997 (rel. 35, Created)			
DT	15-JUL-1998 (rel. 36, Last sequence update)			

DT 30-MAY-2000 (Rel. 39, last annotation update)
 DE Lactadherin (Milk fat globule-EGF factor 8) (MFG-E8) (Sperm
 surface protein SP47) (PP47).
 GN MFG8.
 OS *Scrofa* (P19).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Suidae; Suidae; Sus.
 OX NCBI_TaxID:9623;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Testis;
 RA Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
 RL
 CC -1 FUNCTION: MAY BE INVOLVED IN PHOSPHOLIPID BINDING. ZONA PELLUCIDA-
 BINDING PROTEIN.
 CC -1 SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE PROTEIN.
 CC -1 TISSUE SPECIFICITY: MAMMARY EPITHELIAL CELL SURFACES AND
 SPERMATOZOAN.
 CC -1 SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
 CC -1 SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation.
 CC the European Bioinformatics Institute. There are no restrictions on how
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

EMBL:	Y11683;	CAAT2379.1;	-
DR	HSSP:	P00740;	1BDM.
DR	InterPro:	IPR000561;	EGF-like.
DR	InterPro:	IPR004421;	FA58_C.
DR	Pfam:	PF00008;	EGF_2.
DR	Pfam:	PF00754;	F5_F8_Type-C_2.
DR	SMART:	SM00181;	EGF_2.
DR	SMART:	SM00231;	FA58C_1.
DR	PROSITE:	PS00022;	EGF_1; 2.
DR	PROSITE:	PS01186;	EGF_2; 2.
DR	PROSITE:	PS01285;	FA58C_1; 2.
DR	PROSITE:	PS01286;	FA58C_2; 2.
KW	glycoprotein; Repeat; EGF-like domain.		
FT	DOMAIN	2	41
FT	DOMAIN	91	88
FT	DOMAIN	91	247
FT	DOMAIN	252	409
FT	SITE	67	69
FT	DISULEID	6	17
FT	DISULEID	11	29
FT	DISULEID	31	40
FT	DISULEID	91	247
FT	DISULEID	234	238
FT	DISULEID	252	409
FT	CARBOHYD	41	41
FT	CARBOHYD	372	372
SO	SEQUENCE	409 AA;	45725 MW;
		N-LINKED (GLCNAC...) (POTENTIAL).	
		N-LINKED (GLCNAC...) (POTENTIAL).	
		BOC01AF80029527A CRC64;	

[illegible]

QY 2232 OGRSNANRPOVNNKPKWLOVDFOKTMKYTGTTQGVKSLTSMYKEFLISSDDGHQWT 2291
 DB 312 OGRFNAMTQNTSASISWLDIDJGSKRATGILITOGABDGHGHIQVAAVRAVYGDGWT 371
 QY 2292 LEFONG--KVKVPOGNSDSETPVNSLDPELITRYLRIRHPOSVHQAIALRMEVLGC 2345
 DB 372 EYKDPGASESKIFPGKMDNNSHKKNIFETPOAFVRIQIPVAMHNRITLRFVLLGC 427
 RESULT 14
 MEGM_HUMAN
 ID MEGM_HUMAN STANDARD; PRT: 387 AA.
 AC 008431;
 DT 01-OCT-1996 (Rel. 34, Last Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Lactadherin precursor (Milk fat globule-EGF factor 8) (MFG-E8) (HMEG)
 DE (breast epithelial antigen BA46) (MFGM) [contains: Medin].
 GN MFG8.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominda; Homo.
 OX NCBI_Taxid=9606;
 [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Breast, and Breast carcinoma;
 RA MEDLINE-96313908; PubMed-8639264;
 RA Couto J.R., Taylor M.R., Godwin S.G., Ceriani R.L., Peterson J.A.;
 RT "Cloning and sequence analysis of human breast epithelial antigen
 BA46 reveals an RGD cell adhesion sequence presented on an epidermal
 growth factor-like domain".
 RT DNA Cell Biol. 15:281-286(1996).
 [2]
 RP SEQUENCE OF 170-387 FROM N.A.
 RC TISSUE-Mammary gland;
 RA MEDLINE-91371351; PubMed-190932;
 RA Larocca D., Peterson J.A., Ureia R., Kuniyoshi J., Bistrain A.M.,
 RA Ceriani R.L.;
 RA "A Mr 46,000 human milk fat globule protein that is highly expressed
 in human breast tumors contains factor VIII-like domains".
 RT Cancer Res. 51:4994-4998(1991).
 [3]
 RP PARTIAL SEQUENCE, AND CHARACTERIZATION.
 RC TISSUE-Milk;
 RA MEDLINE-98194924; PubMed-9535276;
 RA Guiffreda M.G., Cavaleto M., Giunta C., Conti A.,
 RA Godovac-Zimmermann J.;
 RT "Isolation and characterization of full and truncated forms of human
 breast carcinoma protein BA46 from human milk fat globule membranes".
 RT J. Protein Chem. 17:143-148(1998).
 [4]
 RP SEQUENCE OF 268-317, AND IDENTIFICATION OF MEDIN.
 RA MEDLINE-99342076; PubMed-10411933;
 RA Haeggevalst B., Naessund J., Sletten K., Westermarck G.T., Mucchiano G.,
 RA Tjernberg L.O., Nordstedt C., Engstrom U., Westermarck P.;
 RT "Medin: an integral fragment of aortic smooth muscle cell-produced
 lactadherin forms the most common human amyloid".
 RT Proc. Natl. Acad. Sci. U.S.A. 96:8669-8674(1999).
 [5]
 RP CHARACTERIZATION.
 RA MEDLINE-97405885; PubMed-9260929;
 RA Taylor M.R., Couto J.R., Scallan C.D., Ceriani R.L., Peterson J.A.;
 RT "Lactadherin (formerly BA46), a membrane-associated glycoprotein
 expressed in human milk and breast carcinomas, promotes Arg-Gly-Asp
 (RGD)-dependent cell adhesion".
 RT DNA Cell Biol. 16:861-869(1997).
 CC -1- FUNCTION: MAY BE INVOLVED IN PHOSPHOLIPID BINDING. BINDS
 SPECIFICALLY TO ROTAVIRUS AND INHIBITS ITS REPLICATION.
 CC -1- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE PROTEIN.
 CC -1- TISSUE SPECIFICITY: MAMMARY EPITHELIAL CELL SURFACES AND AORTIC
 MEDIA. OVEREXPRESSED IN SEVERAL CARCINOMAS.
 CC -1- PTM: MEDIN HAS A RAGGED N-TERMINUS WITH MINOR SPECIES STARTING AT

CC AMINO ACID 264 AND 273
 CC -1- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
 CC -1- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U58516; AAC50549.1; -
 CC EMBL: S56151; AAB1971.1; -
 CC HSSP: P08709; 1BP9.
 CC Genew: HGNC:7036; MFG8.
 CC MIM: 602281.
 CC InterPro: IPR000561; EGF-like.
 CC InterPro: IPR000421; FA58.C.
 CC Pfam: PF00008; EGF, 1.
 CC Pfam: PF00754; F5_F8_Type_C; 2.
 CC SMART: SM0031; FA58C; 2.
 CC SMART: SM0022; EGF_1; 1.
 CC PROSITE: PS00186; FA58C_1; 1.
 CC PROSITE: PS01285; FA58C_2; 1.
 CC PROSITE: PS01286; FA58C_2; 2.
 CC Signal: Glycoprotein; Milk; Repeat; EGF-like domain; Amyloid.
 FT SIGNAL 1 23
 FT CHAIN 24 387 LACTADHERIN.
 FT CHAIN 202 387 LACTADHERIN; SHORT FORM.
 FT CHAIN 268 317 MEDIN.
 FT DOMAIN 24 67 EGF-LIKE.
 FT DOMAIN 70 225 F5/8 TYPE C 1.
 FT DOMAIN 230 367 F5/8 TYPE C 2.
 FT SITE 46 48 CELL ATTACHMENT SITE (POTENTIAL).
 FT DISULFID 27 38 BY SIMILARITY.
 FT DISULFID 32 55 BY SIMILARITY.
 FT DISULFID 57 225 BY SIMILARITY.
 FT DISULFID 70 225 BY SIMILARITY.
 FT DISULFID 212 216 BY SIMILARITY.
 FT DISULFID 387 BY SIMILARITY.
 FT CARBOHYD 238 238 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 325 325 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 329 329 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 350 350 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 387 AA; 43123 MW; 2BE6571DEC83782D CRC64;
 Query Match 4.7%; Score 588; DB 1; Length 387;
 Best Local Similarity 37.3%; Pred. No. 1,1e-23;
 Matches 132; Conservative 69; Mismatches 125; Indels 28; Gaps 9;
 QY 2006 EMLPSKAGIWEVCELIQGEHLAAGNSTFLVYSNKQOTPLGNASGHRIDQITASG----- 2060
 DB 48 DVFPS-----YTCTCKLG---YAGNHC-----EIKGVEPLGEMENINSQIAASVRYTF 95
 QY 2061 -QYQWAPKRLARHYSGSINAW--STEEPSYTKDYLAPMIGITOGAROKFSYXI 2117
 DB 96 LGIDHWPELARINRAGVAVNWPSSNDQPMIYVNLBRMWVGVVTOGASRLASHTEL 155
 QY 2118 SQFIATSLDGRKQNTYGNSTGLAMTFEGVNDSSGKHNIFNPPIARIYLRPHHSI 2177
 DB 156 KAFVVAISLNGHEFD-ETHDNKKKKEFGVNNKNAVNLFEIPEVAQIVRLYPTSCYT 214
 QY 2178 RSTRLMELMGCDLNSCMPLEMSKASIAQITASSYF---TNMEATWSPSKARLHLQG 2233
 DB 215 ACTLRELLGCELCNCAPDLCKNNSIPDQITASSYKTYGHLTF-SNMPSVARLDQOG 273
 QY 2234 RSNAMRPOVNNKPKWLOVDFOKTKRYGVVTOGVASLLTSMYKEFLISSDDGHQWTLF 2293
 DB 274 NFNAWVAGSYGNDQWLOVDLASSKREYVGIITOGARNFVSQFVASYKVAVANDSNAMWTEY 333
 QY 2294 F--QNGKVVYFGNSDSETPVNSLDPELITRYLRIRHPOSVHQAIALRMEVLGC 2345

Wed Jul 2 11:12:45 2003

np_000123.rsp

Page 26

```

07      2257  MKYGVTTQS-----YSLLSMYKVEPLISSDGGHWTLEFQNGV-----VYFGQND 2307
Db      496  KYYKCVIIIGAGRGDSITVEAKAAVKKREKYSJLNGKME-XIDPPRTQAPLFFGNNH 554
07      2308  SFTPVNSLDPLLTLYRLIPQSVNR-TALAMEYICEAD 2349
Db      555  YDPIPIREF-PLPAQVYVPERKSPAGCIGTLEVLGCDWTD 596

```

Search completed: July 2, 2003, 09:43:22
Job time : 58 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OW protein - protein search, using sw model

Run on: July 2, 2003, 09:37:41 ; Search time 138 Seconds
(without alignments)
3510.266 Million cell updates/sec

Title: NP_000123
Perfect score: 12418
Sequence: 1 mgjstcfcflclrlcfca.....wvghalrmeylgceaqdy 2351

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: SP_ARCHA:.*
2: SP_BACTERIA:.*
3: SP_FUNGI:.*
4: SP_HUMAN:.*
5: SP_INVERTEBRATE:.*
6: SP_MAMMAL:.*
7: SP_MHC:.*
8: SP_ORGANELLE:.*
9: SP_PHAGE:.*
10: SP_PLANT:.*
11: SP_RODENT:.*
12: SP_VIRUS:.*
13: SP_VERTEBRATE:.*
14: SP_UNCLASSIFIED:.*
15: SP_VIRUS:.*
16: SP_BACTERIAP:.*
17: SP_ARCHAEP:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	9493	76.4	2343	6 018806	018806 canis fam11
2	9447	76.1	2343	6 062730	062730 canis fam11
3	2819.5	22.7	2119	13 090447	090447 brachydantio
4	2813.5	22.7	2224	4 043737	043737 homo sapien
5	2768.5	22.3	2183	11 088783	088783 mus musculu
6	1382	11.1	1157	11 092024	092024 mus musculu
7	1368	11.0	1157	11 0920H8	0920H8 rattus norv
8	1338	10.8	1158	4 09B057	09B057 homo sapien
9	1333	10.7	1158	4 09C058	09C058 homo sapien
10	1292	10.4	1084	11 09JL97	09JL97 rattus norv
11	1221.5	9.8	1048	6 09XT27	09XT27 ovis aries
12	1104	8.9	216	4 014286	014286 homo sapien
13	1091	8.8	782	4 075659	075659 homo sapien
14	1068.5	8.6	891	4 075180	075180 homo sapien
15	928	7.5	626	13 090772	090772 brachydantio
16	664	5.3	463	11 09R1X9	09R1X9 mus musculu

17	662.5	5.3	480	4 043854	043854 homo sapien
18	661	5.3	426	11 09WTS3	09WTS3 mus musculu
19	655.5	5.3	480	11 035474	035474 mus musculu
20	514.5	4.1	363	6 077718	077718 equus caball
21	488	3.9	335	4 09BTL9	09BTL9 homo sapien
22	469.5	3.8	901	4 09H2E4	09H2E4 homo sapien
23	469.5	3.8	901	4 09H2D5	09H2D5 homo sapien
24	469.5	3.8	906	4 09H2E3	09H2E3 homo sapien
25	469.5	3.8	906	4 09H2D4	09H2D4 homo sapien
26	462.5	3.7	926	11 080217	080217 mus musculu
27	450.5	3.6	921	11 09QX38	09QX38 rattus norv
28	446.5	3.6	919	13 08UVR0	08UVR0 gallus gall
29	446.5	3.6	936	13 08UVQ9	08UVQ9 gallus gall
30	429.5	3.5	609	4 096190	096190 homo sapien
31	429.5	3.5	644	4 0961H5	0961H5 homo sapien
32	429.5	3.5	704	4 09H2E1	09H2E1 homo sapien
33	429.5	3.5	779	4 09NTT3	09NTT3 homo sapien
34	415.5	3.3	923	13 08ORX6	08ORX6 brachydantio
35	411.5	3.3	858	5 076470	076470 lytechinus
36	389.5	3.1	555	4 09H2E2	09H2E2 homo sapien
37	308.5	2.5	775	4 096PD2	096PD2 homo sapien
38	308.5	2.5	775	4 08TDX2	08TDX2 homo sapien
39	299.5	2.4	769	11 0912V3	0912V3 rattus norv
40	295	2.4	769	11 0912V3	0912V3 mus musculu
41	273.5	2.2	2647	5 09U4X0	09U4X0 plasmodium
42	273	2.2	51	4 09U005	09U005 homo sapien
43	271	2.2	1128	11 088442	088442 mus musculu
44	269	2.1	70	6 097899	097899 atelapha beliz
45	266	2.1	764	11 09DZL5	09DZL5 mus musculu

ALIGNMENTS

RESULT 1	PRELIMINARY:	PRT:	2343 AA.
018806			
AC	018806:		
DT	01-JAN-1998 (TREMBLrel. 05, Created)		
DT	01-JAN-1998 (TREMBLrel. 05, Last sequence update)		
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)		
DE	Factor VIII.		
GN	F8.		
OS	Canis familiaris (Dog).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.		
OX	NCBI_TaxID=9615;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=LIVER;		
RA	Cameron C., Nolley C., Hoyle S., McElvyn L., Hough C., Kamlsue S.,		
RA	Giles A., Lillierap D.;		
RT	"The canine factor VIII cDNA and 5' flanking sequence."		
RL	Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.		
CC	-1 SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.		
DR	EMBL; AF016234; AB87412.1; -		
DR	HSSD; P00451; ICFG.		
DR	Interpro: IPR001117; Cu-oxidase.		
DR	Interpro: IPR000421; FAS8_C.		
DR	Pfam: PF00394; Cu-oxidase; 3.		
DR	Pfam: PF00754; F5_P8_Type_C; 2.		
DR	SMART; SM00231; FAS8C; 2.		
DR	PROSITE; PS01285; FAS8C_1; 2.		
DR	PROSITE; PS01286; FAS8C_2; 2.		
DR	PROSITE; PS00079; MULTICOPEP_OXIDASE1; 3.		
DR	SEQUENCE 2343 AA; 265829 MW; AB54FAE571C3B399 CRC64;		
SO	SEQUENCE		
Query Match	76.4%; Score 9493; DB 6; Length 2343;		
Best local similarity	77.1%; Pred. No. 0;		
Matches 1817; Conservative 196; Mismatches 319; Indels 24; Gaps 12;			
1 MQJLSLFCFLCRLRCFSARRRYLVGLAVLSMDYQSD-IGELPYDKRPPRPSPFP 59			

Db 1 MQVELYTCFCFLCLPFSLSATRKYYLGAVELSWDYQSDLLSALHADTSEFSRPSGLD 60
Qy 60 NTSVYKKTLFVEFDHLEFNIAKPRPMWGLGPTIOAEVDTVYTLKNASHVSLA 119
Db 61 TTSVTRKTVFVEFDHLEFNIAKPRPMWGLGPTIOAEVDTVYTLKNASHVSLA 120
Qy 120 VGVSYWKASGAEAYDDOTSOREKEDKVPFGSGSHYVWQVLEKNGPMA5DCLTYSYL 179
Db 121 VGVSYWKASGAEAYDDOTSOREKEDKVPFGSGSHYVWQVLEKNGPMA5DCLTYSYL 180
Qy 180 HDVLKVDLNSGLIGALLVCREGSLAKETQTLKFFLEFAVEDEKSMHSETKSLMOOR 239
Db 181 HDVLKVDLNSGLIGALLVCREGSLAKETQTLKFFLEFAVEDEKSMHSETKSLMOOR 238
Qy 240 DAASAPAKMHTVNGVYNSRLPOLICCHRSYVWYVWQVLEKNGPMA5DCLTYSYL 299
Db 239 ---AEOHEHHTINGVYNSRLPOLICCHRSYVWYVWQVLEKNGPMA5DCLTYSYL 294
Qy 300 HROASLEISPIITELAQOTLMDLGOTLFCIHSHOHDMGMEAYVVDSCPEPOLRKN 359
Db 295 HROASLEISPIITELAQOTLMDLGOTLFCIHSHOHDMGMEAYVVDSCPEPOLRKN 354
Qy 360 EEAEDYDDOTSEMDVYRFDNDSPFOTRSVAKKHPTVHYIAEEDMDYAPLV 419
Db 355 ED-KDIDGGLSDMDVYSFDSDSPFOTRSVAKKHPTVHYIAEEDMDYAPLV 413
Qy 420 APDRSYKSOYLNGPOTIRGRYKRVPMAYDTETKREAIQHSGLIGLVYEGDT 479
Db 414 TPNDRSHKNTLYNNGPOTIRGRYKRVPMAYDTETKREAIQHSGLIGLVYEGDT 473
Qy 480 LLIIFKNQASRPNTYPIGTDVPRPLYSRRLPGVYHLDLDFLIGELIFKKMTVYEDG 539
Db 474 LLIIFKNQASRPNTYPIGTDVPRPLYSRRLPGVYHLDLDFLIGELIFKKMTVYEDG 533
Qy 540 PTKSDPCLTRYYSFVNNERDLASGLIGLVYEDGQNGOIMSDKRVYLFYED 599
Db 534 PTKSDPCLTRYYSFVNNERDLASGLIGLVYEDGQNGOIMSDKRVYLFYED 593
Qy 600 ENRSWYLTENIOERFLPNPAGVLEDEPFOASINMHSINGVPSOLSYCHAEVAYYL 659
Db 594 ENRSWYLTENIOERFLPNPAGVLEDEPFOASINMHSINGVPSOLSYCHAEVAYYL 653
Qy 660 SIGAQDTFLSVFSGYTFKHKMYEDTLTFPFGSETVMSMNPDLWGLGHSNDFNR 713
Db 654 SIGAQDTFLSVFSGYTFKHKMYEDTLTFPFGSETVMSMNPDLWGLGHSNDFNR 713
Qy 720 GMTALTKVSSCDKNTDYEDYEDISAYLLSKNNAIEPRSEFSONSRHPTROKOFNAT 779
Db 714 GMTALTKVSSCDKNTDYEDYEDISAYLLSKNNAIEPRSEFSONSRHPTROKOFNAT 773
Qy 780 IPENDIKTDPFAHKTMPKIQONSSDMLMLROSPFHGLSLDQOAKYEFSDP 839
Db 774 IPENDIKTDPFAHKTMPKIQONSSDMLMLROSPFHGLSLDQOAKYEFSDP 831
Qy 840 SPGAIDSNNSLSMTFPRPOLHSGDMVTFPESGLOLRNEKLGTTAATTELKLDKFKVS 899
Db 832 SPGAIDSNNSLSMTFPRPOLHSGDMVTFPESGLOLRNEKLGTTAATTELKLDKFKVS 891
Qy 900 TSNNLIS--TIPSDNLAAGTONTSLGPPMPVYVDSQDPTLFGKSSPPLTESGPIST 957
Db 892 TSNNLIS--TIPSDNLAAGTONTSLGPPMPVYVDSQDPTLFGKSSPPLTESGPIST 951
Qy 958 SEENDSKLLESGLIANSOESSWGNKVSSTESGRLFKGRAHGPALLJKNALFKVYSIL 1017
Db 952 SEENDSKLLESGLIANSOESSWGNKVSSTESGRLFKGRAHGPALLJKNALFKVYSIL 1011
Qy 1018 KTNTSNNSATNRKTHIDGSLILNPSVWONT--LESDFEFKVPJLJHDRLDKNAT 1076
Db 1012 KTNTSNNSATNRKTHIDGSLILNPSVWONT--LESDFEFKVPJLJHDRLDKNAT 1071
Qy 1077 ALRLHNSNKTTSKNNENVOOKKEGPIPPDAPNDMSFFKMLFLPESARWIOPTHGNS 1136
Db 1072 ALRLHNSNKTTSKNNENVOOKKEGPIPPDAPNDMSFFKMLFLPESARWIOPTHGNS 1127

Qy 1137 LNSGQSPKOLVSLGPERKSVGEQNTLSSEKKNVYVKGGEFTKDVLEKVEPSSKNLELT 1196
Db 1128 LNSGQSPKOLVSLGPERKSVGEQNTLSSEKKNVYVKGGEFTKDVLEKVEPSSKNLELT 1185
Qy 1197 NLDNLHENTHNOEKTIOEIEKKEKTLLOENVVLPOLHNTGKKNKMLFLSTROVE 1256
Db 1186 NLDNLHENTHNOEKTIOEIEKKEKTLLOENVVLPOLHNTGKKNKMLFLSTROVE 1245
Qy 1257 GSYDGAAPVLQDFRSLNDSTNRKTHAHFSKKEEENLEGLQNTKOIYKACTRI 1316
Db 1246 GSYDGAAPVLQDFRSLNDSTNRKTHAHFSKKEEENLEGLQNTKOIYKACTRI 1305
Qy 1317 SPNTSOONFVTOSSKRALKOPFLPEETELEKRTIYDOSTQSKNNKMLFLSTLOIDY 1376
Db 1306 SPNTSOONFVTOSSKRALKOPFLPEETELEKRTIYDOSTQSKNNKMLFLSTLOIDY 1364
Qy 1377 NEKEGATIOSPLSDCLTSHSIPQANSPLPIAKVSFSPSTIRYTLRVLFDONSHLP 1436
Db 1365 NEKEGATIOSPLSDCLTSHSIPQANSPLPIAKVSFSPSTIRYTLRVLFDONSHLP 1424
Qy 1437 AA---SYRKDSGVGSHFLQGAKNLSTALITLMTGDQREVSLGTSATNSVYK 1492
Db 1425 AA---SYRKDSGVGSHFLQGAKNLSTALITLMTGDQREVSLGTSATNSVYK 1484
Qy 1493 KYENVLKPOLPKTSKGVYLLPKYHIOKOLPPTENSGSPGHLVLGSLDGEAI 1552
Db 1485 KYENVLKPOLPKTSKGVYLLPKYHIOKOLPPTENSGSPGHLVLGSLDGEAI 1544
Qy 1553 KNEANRPGKVPFLVATESAKTPSKLLDPLAMDNHYQIOPREKMSQESPEKAFK 1612
Db 1545 KNEANRPGKVPFLVATESAKTPSKLLDPLAMDNHYQIOPREKMSQESPEKAFK 1604
Qy 1613 KDOTLSLACESNHAIAINEGONKPEIEVMAAGTBEGLGSONPVYKROREIRT 1672
Db 1605 KDOTLSLACESNHAIAINEGONKPEIEVMAAGTBEGLGSONPVYKROREIRT 1664
Qy 1673 TLQSDQEEIDYDTIYEMKEDFDIYDENQSPRSFOKTRHYTAAVERLMDGNS 1732
Db 1665 TLQSDQEEIDYDTIYEMKEDFDIYDENQSPRSFOKTRHYTAAVERLMDGNS 1724
Qy 1733 SPHYLRNROSGSVPOKRVYPOEFTDGSOTPLYRGELNHLGLAGPYRAVEEDNIY 1792
Db 1725 SPHYLRNROSGSVPOKRVYPOEFTDGSOTPLYRGELNHLGLAGPYRAVEEDNIY 1784
Qy 1793 TFRNOASRPYSFYSLSIYEDDQGAEPKRVKPNETKYTFKVOHMAPTKDEPDK 1852
Db 1785 TFRNOASRPYSFYSLSIYEDDQGAEPKRVKPNETKYTFKVOHMAPTKDEPDK 1844
Qy 1853 ANAYPSDVLKDVHSGILGPLVYCHNTLNPAGROYTVOEFALFTIYDEKSWYTE 1912
Db 1845 ANAYPSDVLKDVHSGILGPLVYCHNTLNPAGROYTVOEFALFTIYDEKSWYTE 1904
Qy 1913 NNERCRAPCNIOMEDPTFKENTRFAHNGYIMOTLRLGVAAOQORIRKWLMSGNEI 1972
Db 1905 NNERCRAPCNIOMEDPTFKENTRFAHNGYIMOTLRLGVAAOQORIRKWLMSGNEI 1964
Qy 1973 HSIHSGHYVYRKKKEEKYALNYLYPGVFEVEMLPKAGIMVEECLIGEHLAAGMTL 2032
Db 1965 HSIHSGHYVYRKKKEEKYALNYLYPGVFEVEMLPKAGIMVEECLIGEHLAAGMTL 2024
Qy 2033 FLVYSSKQOTPLGMA5GHIIDPOTASGOYGOMAKRLARLYSSISINAMSTKXPFMIV 2092
Db 2025 FLVYSSKQOTPLGMA5GHIIDPOTASGOYGOMAKRLARLYSSISINAMSTKXPFMIV 2084
Qy 2093 DLAPMIIHIGITGAROKSSLYISOFIINTSLDGKRMQYTRGNSTGLTMAVFGVND 2152
Db 2085 DLAPMIIHIGITGAROKSSLYISOFIINTSLDGKRMQYTRGNSTGLTMAVFGVND 2144
Qy 2153 GIKHNTFNPITIAITRIHPTPHYSIRSTLAMELMDGDLNSGMPLGMSKASIDAOITAS 2212
Db 2145 GIKHNTFNPITIAITRIHPTPHYSIRSTLAMELMDGDLNSGMPLGMSKASIDAOITAS 2204

[illegible]

RESULT 2	
062730	
ID 062730	PRELIMINARY;
MC 062730.	PRT; 2343 AA

DT 01-~~AUG~~-1998 (TREMblrel. 07, Created)
 DT 01-~~AUG~~-1998 (TREMblrel. 07, last sequence update)
 DT 01-~~DEC~~-2001 (TREMblrel. 19, last annotation update)
 DT Factor VIII.
 OS *Canis familiaris* (Dog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
 OC NCBI_TaxId=9615;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=KIDNEY, AND SPLEEN;
 RC Gordy P.W., Bowen R.A.;
 RA "Characterization of the canine factor VIII cDNA."
 RT Submitted (FEB-1993) to the EMBL/GenBank/DBJ databases.
 RL 1. SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
 EMBL: AF009489; AF005384.1; -.
 DR HSSP: R00451.1; CDF.
 DR InterPro: IPR001117; Cu-oxidase.
 DR InterPro: IPR000421; FAS5_C.
 DR Pfam: PF00394; Cu-oxidase; 3.
 DR Pfam: PF00754; F5_p8_type_C; 2.
 DR SMART: SM00231; FAS5C_1; 2.
 DR PROSITE: PS001265; FAS5C_1; 2.
 DR PROSITE: PS001285; FAS5C_1; 2.
 DR PROSITE: PS00079; MULTICOPPER_OXIDASE1; 3.
 DR SOURCE 2343 AA; 263613 MW; F612D744DAAD99 CRC64;
 QO

Query Match	76.18;	Score 9447;	DB 6;	Length 2343;
Best Local Similarity	76.88;	Pred. No. 0;		
Matches 1809;	Conservative 199;	Mismatches 324;	Indels 24;	Gaps 12;

Qy	1	MOIELSTCEPLCLARCCEASAPRXYIAYELSMYDNOCD - LGELVYAPARPPRPYKSPFP	59
Qy	1	MOVELYTCCLCLPSPSLATKRIYIGLAVELSMYDNOCDLSALHADJSEFSRPSDLP	60
Db	60	NTSVYAKTLEVEETDHLFNIAKPRPMMGILSPYIOAEVYDYVYITLKNAASHVSLHA	119
Db	61	TTSVYAKRYEVEETDLEFNIAKPRPMMGILSPYIOAEVYDYVYITLKNAASHVSLHA	120
Qy	120	VGVSYWMASSGALYEDDOTGOREKEDDYKPPGGSHYVVOYLKENGPMASDPLCLTYSYS	179
Db	121	VGVSYWMASSGALYEDQTSQREKEDDNYIGEBSHYVVOYLKENGPMASDPLCLTYSYS	180
Qy	180	HYDLYKDLNSGLIGALLVCRRGSTAKERTQTLKFLFLPAYDDECKSMHSETKNSLMODR	239
Db	181	HYDLYKDLNSGLIGALLVCRRGSTAKERTQTLDEFLFLPAYDDECKSMHSETKNSLTO --	238
Qy	240	DAASARAPKQHTVNGVYKRSILPGLIGCHRSYVWYIOMGTTPEVHSIEFLGHTLYEN	299
Db	239	---AEAOHETHTINGYVYKRSILPGLIYCHRSYVWYIOMGTTPEVHSIEFLGHTLYEN	294
Qy	300	HROASIEIISPTTLTQOTLLMDLGOFLLECHISSHOHOMEMAYUVYVDSCEPDLRYKNN	359
Db	295	HROASIEIISPTTLTQOTLLMDLGOFLLECHITSHOHOMEMAYVVDSCPEPDLRYKNN	354
Qy	360	EEAEEDYDDLLTDEMDVFRPDDNSPSFIQIRSVAKKRRKTVVHYIAAEEDMDYAPVLV	419

Dh	355	BD - KDYDUGLXGSDMYVSPDDSSPFQIRSVAKHKKPMVHYIAAEEBMDXAPSGP	413
Oy	420	APDRASYSOYLXNGNORIGRRKRYKRYEMAYTDEPRTKREAIOMEHESLIGLPLLYGVGDT	479
Dh	414	TPNDRSHKNTLYLNGRQORIGKRYKRYEAVATDEYTKTREALOTRESGLDPLLYGEVGT	473
Oy	480	LLIIFKNQASRPNIYPHGIDTVRPLYSRRLPGVAKHLKDPPLILGEIFYKWTYVEBG	539
Dh	474	LLIIFFKQASRPNIYPHGINTYPTLHTGRLPGVAKHLKDMPLILGEIFYKWTYVEBG	533
Oy	540	PTKSDPCLLRYRYSSEFVNNERNDLASGLIGLILCYEESYDORGNQIMSDRNNYLFEYVD	599
Dh	534	PTKSDPCLLRYRYSSEFINLERDLASGLIGLILCYEESYDORGNQIMSDRNNYLFEYVD	593
Oy	600	ENRSWYLTENIOEFLPNPAGVOLEDPEFOASNTMHSINCYVDSLOLSTCLHEVAWTYL	659
Dh	594	ENRSMYLTEDMOEFLPNADYVQPHDPEFOLSNTMHSINCYVDSLOLSTCLHEVAWTYL	653
Oy	660	SIGQOTPLSEYFSGTTFYKHKMYEDOTLTFPPSGEYVMSENPGLMILGCSNPSFRK	719
Dh	654	SYGQOTPLSEYFSGTTFYKHKMYEDOTLTFPPSGEYVMSENPGLMILGCSNPSFRK	713
Oy	720	GMYALLVSSCDKNTDYEDYSIEDISAYLLSKNNAIEBRSESONSRHPSTROKOPNAT	779
Dh	714	GMYALLVSSCMMNDYEDYEDIEPPLNNNNYIKRSPSONSRHPSTYKEROUKATT	773
Oy	780	IPENDIKETPMFAHRTPMPTKIONVSSDDLMLRQSPPHGLSISDLOEAKETPSDOP	839
Dh	774	TPENDIEKIDLOGEFRTOLIKQASVSSDDLMLLQGNPRLGFLSTDLREATOR--ADHG	831
Oy	840	SPGAISSNLSLSEMTFRQLHSGDMVTFPPSGLOLRLNKLGTATATEKTLKDPKVS	899
Dh	832	SMGAIENKNGPPEVSLARELHSHSDREPRPEEJOLRLNKLGNITTYELKTLKLKISS	891
Oy	900	TSMNLIIS--TIIPBDNLAACTDWTSLGPSPMPVPHYOSQDITTLREKSSPLTSSGGPFL	957
Dh	892	SDSLMSTPLIIPDKLAAKTEKTSIGLCPKPNVSHFGKGLGTYEPNNSHSLDOSVPLL	951
Oy	956	SEENKDSLLESGJMSOESSMGNKVNSTESGRLFKGRAPALLTKDNALFEVSISSL	1017
Dh	952	SEEDNDSLLLEAPLMIQIOESSLRENVJLSEMSNNLFEKERRIGCPALIKDNALFEVSISSV	1011
Oy	1018	KTNKTSNNKSNATNKRTIIDPSLLEIENSYSWQNI--LESOTERKATYPLIHDIMAKMNT	1076
Dh	1012	KTNRPAYVLTNKNKRTIYVAPLLIENSYSWODIMLERTERKATYSLIHNETFMDRNTT	1071
Oy	1077	ALRLHNSNTTSSKMKMEVQOKGCPRPDQANPMSPEFKLPLPSABWORTGKRS	1136
Dh	1072	ALGLNHSNNTTSLSKVNEAHOKKEDVPLRAENPDLSSKTLPELPD--WIKTKGKS	1122
Oy	1137	INSGGSPSKQVLSLQPEKSVBGONFTLSEKNKRYVNGEFTYDGLKEMVPPSSRLYLT	1196
Dh	1128	LSSEBRPSPQOLSTLSEKSVKQDNQFSEB-KVAYEDEFETKDELDE-IFPNKKSIFPA	1185
Oy	1197	NLDNLHENNTHNOEKTIOBEIKKKTLLIOENVVLPOLIHVTYTKNFKMKTLLSTQONE	1256
Dh	1186	NLANVOENDTYNOEKKSLIEIERKEKTLQOEVAVLPDAHMTIKTNKFLNLFLLSTQONRA	1245
Oy	1257	GSYDAVAPYLOPFRSLNOSTWRTKRTKHAHSSKGEEMLELGNOTKQIYEKUYACTRTI	1316
Dh	1246	GLEQPYTPILLODTRSLNDSHSEGIHMANFSKIREANLBEIGNOTNOMVERFPSTTR	1305
Oy	1317	SPNTSOONFTYOSKRALKOPLRPLELEETELEKELIYDOSTOMSKNNKHLPLSTLOYD	1376
Dh	1306	SSNMS-QHVVITQOKRSIKQPLRSOGELKEPKRYANDSTOMSKNNKMYLAOGTLQIYER	1366
Oy	1377	NEKEGATQSPISDCLTSHSITPDANRSPPLAKYSSFPSTIRPLIYLRVLPDQNSGSH	1433
Dh	1365	NEKEKRATQSPISDSCSMNHVTIOMDMSALPAVKESASPSYRRTROLKTLPSOHNSSH	1422
Oy	1437	AA---STYRKSDGVOESSHFLQAKKNNKSLIILTLLENTGQOREVGLSGTASNVYTK	1492

Wed Jul 2 11:12:46 2003

np_000123.rspt

Page 4

Dh	1425	ASACNTYFRFRRTSGVDESHFLDQAKRNNLSLAFTVLTGITGQCKFSGSKSATNPQYK	1484
Qy	1493	KVENTVYLPKRDPLDKTSGVELLPVYHLYOKDLPETSGNSPGHLDIYESLLOGBEAI	1582
Dh	1485	KLENTVYLLQGLSEFSDKVELLSQYHNDQDSDSPFKTSSNDSPGHLDMGKIFLOKTQGPV	1544
Qy	1553	KWNEANRPQVPLPKVLAIESKAPKPSKLLDPLAMDNHGTQIPREEEKSQEKSPEKTAfk	1612
Dh	1545	KMNNTSPGVVPLPKVAETSESKIPSLKGLVLAAMDNDYDIPPEEKSKOKSQTWYfK	1604
Qy	1613	KKDTLLSLANCSNNAIAINEGONKPELEVTWAKOGTERLCSQNPVYLKRHORETRT	1672
Dh	1605	KKDTLLSLGECENNSTAINEGODKFORAMMAKQGERGLCSQNPVYSKHOREIVYT	1664
Qy	1673	TLDQDELEIDYDQTSIVSEKKEDDIDIEDENOSPFOSFKTRFYFAAVERLMQGMS	1732
Dh	1665	TLDPEEDKPEYDDTSPTEKKREDPDIDGYEDQGLRSFKQKTRHYFAAVERLMQGMSR	1724
Qy	1733	SPHLNRARQSGVPOFKKYVFOEFTGSTQPLVYGLNELHLLDLPYRAVEDNTVY	1792
Dh	1725	SPHLNRARQSGVPOFKKYVFOEFTGSTQPLVYGLNELHLLDLPYRAVEDNTVY	1784
Qy	1793	TFRQNASRPSYFSYSLISTEEDQOGAPRKPKVFNKTYFKVQNHAAPIKDEPCK	1852
Dh	1785	TFRQNASRPSYFSYSLISTEEDQOGAPRKPKVFNKTYFKVQNHAAPIKDEPCK	1844
Qy	1853	AMAFSDVDLEKRYHSGLLGIPLVCHNTLNPAGHQVTVQEFALFTPIDETKSWYfE	1912
Dh	1845	AMAFSDVDLEKRYHSGLLGIPLVCHNTLNPAGHQVTVQEFALFTPIDETKSWYfE	1904
Qy	1913	NMERNCRAPCNIDMEPTKENVFPAHINCYIMDTLPGLVAAQOQIRIWTYLLSMGSENI	1972
Dh	1905	NMERNCRAPCNQVKEDPTLKEFNFPAHINCYVDTLPGLVAAQOQIRIWTYLLSMGSENI	1964
Qy	1973	HSIHSGHVFYTKKREYKFMALYNYLYGVEFYEMLPKSAQIMRECELGIEHILHAGMSTL	2032
Dh	1965	HSIHSGHVFYTKKREYKFMALYNYLYGVEFYEMLPKSAQIMRECELGIEHILHAGMSTL	2024
Qy	2033	FLYVSNKQCPRLMASGHLRDPQTITASQGVQGMAPLALAHSGSINAMSKREPMITV	2092
Dh	2025	FLYVSNKQCPRLMASGHLRDPQTITASQGVQGMAPLALAHSGSINAMSKREPMITV	2084
Qy	2093	DLLAPMIHGIKTQAGKQKSFSSLYISQPIIMYSLDGRKKQYTRGNSGTGLMVEFGVDS	2152
Dh	2085	DLLAPMIHGIKTQAGKQKSFSSLYISQPIIMYSLDGRKKQYTRGNSGTGLMVEFGVDS	2144
Qy	2153	GIKHNIENPPIIARYRLRPTHTYSIRSTLRMELMGCDLSCSMPLGMSKAISDAQITAS	2212
Dh	2145	GIKHNIENPPIIARYRLRPTHTYSIRSTLRMELMGCDLSCSMPLGMSKAISDAQITAS	2204
Qy	2213	SYFTNMFATSPSKARHLHOGSNMAMPQVNNPKEMLOYDPKTKWYGVYGTTOGVSLAT	2272
Dh	2205	SYFTNMFATSPSKARHLHOGSNMAMPQVNNPKEMLOYDPKTKWYGVYGTTOGVSLAT	2264
Qy	2273	SKTYAEFLISSQDGQHTLFPQNKAKVYQGNQDSFTYVNSLDPLRLTRYLRHPOSM	2332
Dh	2265	SKTYAEFLISSQDGQHTLFPQNKAKVYQGNQDSFTYVNSLDPLRLTRYLRHPOSM	2324
Qy	2333	VHQIALLRMVLTGCEAQ 2348	
Dh	2325	AHHIALRLVLTGCDTQ 2340	
RESULT 3			
Q90X47			
AC	Q90X47	PRELIMINARY;	PRT: 2119 AA.
DT	01-DEC-2001 (TRENDArel. 19, Created)		
DT	01-MAR-2002 (TRENDArel. 20, Last sequence update)		
DE	SC:hz2015.3 (novel protein similar to vertebrate coagulation factor V and VIII).		
GN	SC:hz2015.3.		

OS *Brachyotona reirio* (Zebratfish) (*Zebrá danio*).
 OS *Euhayotoa*: Metazoa, Chordata: Craniata: Vertebrata: Euteleostomi,
 OC Actinopterygii: Neopterygii, Teleostei, Ostariophysi, Cypriniformes,
 OC Cyprinidae: Danio.
 CX NCBI_TaxID=7955;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA LLOYD D.;
 RL Submitted (OCr-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AL590146; CAC94896.1; -.
 DR InterPro: IPR001117; Cu-oxidase.
 DR InterPro: IPR000421; FA58_C.
 DR Pfam: PF00394; Cu-oxidase; 2.
 DR Pfam: PF00754; F5_F8_type_C; 2.
 DR PROSITE: PS01285; FA58C_1; UNKNOWN 2.
 DR PROSITE: PS01286; FA58C_2; UNKNOWN 1.
 DR PROSITE: PS00079; MULTICOPPR_OXIDASE1; UNKNOWN 3.
 SQ SEQUENCE 2119 AA; 240643 MW; DCDEB06FFA8761E6 CRC64;

Query Match	22.78;	Score 2819.5;	DB 13;	Length 2119;
Best Local Similarity	30.28;	Pred. No. 1.3e-155;		
Matches 751;	Conservative 379;	Mismatches 826;	Indels 529;	Gaps 68;

```

0Y 11 LCLLRF -CFNST---RRYKAGVLELSDMQDOLGELBVPDAFRPRYKSPFRPNTSVYK 66
0Y 13 LMLLNFLOCHHCHAVEHRYTIANVINDYTSQGO-----KIOGSK 53
0Y 67 KTLVEFTDLFLFIARPRPMMGLIGPTLQAEVYDTVITLKNMASHPVLAVGVSTK 126
0Y 54 KYVYAEYVEG- FQOPAHPLSSGLLGRKQGGEDITITFENMADHCSLPHGATYK 112
0Y 127 ASEGAEVYDDQTSQREKEDKVPFGGSHTYVWYKENGMASPDLCLTYSYLSHVDYKD 186
0Y 113 OSEGSEYFDNYSLEKENDVIOGGEHTYOMDYTSVPTLADPRCITITVYLSHPIDYD 172
0Y 187 LMSGLIGLALVCREGSLAKERTQTLH---KFILFVPEGKSMSETKMSLQDDAAS 243
0Y 173 YMTGLIGPLMCKKGTLLDSSNQ -IHFQESVTLFVPEKNSVTS-----TGOS 221
0Y 244 AAMPMKMTYVGVYNSLJGLIGCHRSYVHYVIGAGTPEYHSFLEGHPTLYRNHROA 303
0Y 222 PQLPWAYKTITNGTNSVDDLICAHRSYMHLLGMSSEPLFSVHFNGOYLHDDHKTS 281
0Y 304 SEISPIPTTELTAGTLLMDLQGFLECHLSSHQDGMKAEATYKVDSPF-EPOLRMKNNE 361
0Y 282 AGILISGATSAISATGHHGGRWLVSHTLSKLEGLAGLTLNKCDETAPAKRRLTID- 340
0Y 362 ABDYDDDLTUSEDDYVAFDDNPSFSIQIRSAKKHKPTWYHIAEEDMDYALVTLAP 421
0Y 341 -----KKESQEMTYMAAEVYIMDYPAMPMPEN 367
0Y 422 DDRSKYQYLTNGPORGIRKRYKKRFMAYTDEFPKR---EALIQHSESLIGLPLLYGEVD 478
0Y 368 MGDPRSKYTLQSGPQIRKAKKVAFTQYKQDMKERAEDNQKRELTALIGLAPYLAQID 427
0Y 479 TLLIFKNOASRPVNIPIGIDVRLPSRLPBGVYHNLKDPILIGEIFYKMYTIVD 538
0Y 428 IRIYFKKAKSRPISILPHOITIDKAAAGASYPOGCH-QYTSVPGEGTITTYTSTVED 485
0Y 539 GPTKSDPCLCTRYYSSFVNNERDLASGLIGPLLCIKESYDORGMOJMSDKRNYLFSVF 598
0Y 486 VFLPDSDPCLCTRMYHSAVAPRDIASGLVGLLCKSQSLNKNVOLLAKDEOHAMEYTF 545
0Y 599 DENRSMYLTENIOFLELPNAGVOLADPEFOASNIHMSINGVYFSLQ-LSYCLHEVAAYV 657
0Y 546 DENKSMYODENINITYCSDPRKYKADDPFYKSNMHTINGVYEGSGELGCHBEIATYWH 605
0Y 658 ILSIGKORDFLSVESFSTPKRMKMYEDTLTLPFGESEVYFVMSNPLATLIGCHSNDR 717
0Y 606 VSVSGEDQYIOTAFYCHHPTELKNREDDILSLFPMNGERTITMNMNATISWLASNSHDS 665
0Y 718 NRGKATLLKAVSCDKANTGDY-JEDSYDLSIYLSLKNNAITERRPSQSQRNPSTROKQEN 776

```

```

Db 666 TKGARVFKDLECFR---DIVIEYED-----GKTAKEPPT-----701
Qy 777 ATTIPENDIEKTDPMFAHRTMPKIONVSSDMLMLROSPFHGLSLDQEAERTFS 836
Db 702 -----NEIKKEEPRAR-----713
Qy 837 DDPSPGAIDSNLSSEMTHEFAPLHSHGDMVETPESGLQLRLANEKLGATTAELKLDK 896
Db 714 ----PPVDEYSDLEPAET-----LNIRFNNVKDEVEIDLDL 746
Qy 897 VSSSTNNLSTIPSDNLAAGTDMTSSLAGPSPMVHYDSQDITLTFGKKSPLTESGGSL 956
Db 747 FLDDODGDLPIVEEKSLOSSNEN-----LANATQSTIEHNG-----783
Qy 957 LSEBNDKSLLESGLANSOESSMCKNVSTESGRLPKGRAGHPALLTKONALFVYSLSL 1016
Db 784 -----LMEEGDLDGESSNMVINDSTDK-----AL 809
Qy 1017 LKTNKTSSNNS--ATNRKTH--IDGPSL--DIENSPSWONILESDTEFEKVTPLIDR 1068
Db 810 LETTTTFDSNRVVALNMTDSTILDEPIVERKVRASAPSKPMNESESVTAKNFTEH1-- 866
Qy 1069 MLADKNAFALRLNHSNKTTSKMMEMVOOKKECP1PP-DAQNPDMSF-----FKMLFLP 1122
Db 867 ----NSLERINAIPTETINNTWTETHTDFTSPDGSTGEMKFTLEDDTALLMS 921
Qy 1123 ESARMIQRTGKNSLNSGSPSPQVLVSLGPEKSEVQGNF-----LSEKNVYVGK 1173
Qy 922 ESEPIOSNONSBNRIAPQELMAKDGVDVSNNSKNOIFKYNVPSGDTLSNKSQIY-- 979
Qy 1174 GEFTKDVGKLEMPFPSSNLFNLNMLHENTHNOEKIOEIEKEKTLIOBNVLPQI 1233
Db 980 -----EDFVLLDSSYFSEMSST--TMEYDVOSKQVYKGSKEKTAOS 1018
Qy 1234 HVTGTRKNFMKMLFLSTRONVEGYDAVAPYLD-----FRSLN--DS 1276
Db 1019 QELSTRTKYSGEIIIESLPITSAFNLSSSYLNRNLSSESSNETLEMSSNATFSS 1078
Qy 1277 TNRTKHT-----AHFSKKG--EENELGCONOTQIVEKVACTRIPIPTSOONFOTORS 1330
Db 1079 TNATSSCSGATFADFSTWTFPSNATFSDPSNRISQMSDSSNATLSDSSNAT--LSSS 1124
Qy 1331 KRALKQFLRLPETELEKRIIYDDTSTQMSKMKHLTPS-----TLQIDYN 1377
Db 1135 NATLSD-----SSNATLSDSSNATLSSNATFSDSSNATFSDSSNATLFGVYS 1184
Qy 1378 -----EKEGATIQSPISDCLTRSHSIPQANRSPPLIARVSSFSIRPIYLRVL 1427
Db 1185 SNTTLDSPLESEEMTEYTLSSANDTI-KSHSEVYNTSOL-----SSSETENISL--L 1235
Qy 1428 FQDNSSHLPASYSRRKDSGVESHFLQAGKKNLTLA1LTLEMTGQORVSGISGTSATN 1487
Db 1236 YGSLN-----ASSMKNDSESESEEEVYIYNKNH--SEALITSHL--DOKE--EHMGYESKH 1286
Qy 1488 SVTYKKVENVV-----LPRPDLPTSGKV--ELLPRVH1-----YOKDLEPTET 1529
Db 1287 ELVHKELPDHMKYVKDKSAANSKPKIREKKKQYQVAKRKGYKMTKKSADVKPQR 1346
Qy 1530 SNGSPGHLDVESGLLOTBEGATIKMNEANPGCVFLVLAFTSS--AKTPSKILDLAMD 1588
Db 1347 SSFSP-----RQFGSVLTPRGSRP-----VSEBDELTEKRPVIVIVPRRDFN 1388
Qy 1589 HYGQIQPKEMKSOEKSEKTAFAKKDTIISLACESHNAIAINSONKPELEJETAQO 1648
Db 1389 DYELIYIPKODEAD-----FDGLL-----DHPE-----1411
Qy 1649 GRTERLCSNPYLARKHOREITRTTLOSDEIDYDITSEMKEDFDIYD--EDENOSP 1707
Db 1412 -----EYEVKEKDYD--KTADVQALDTSQHLK 1440
Qy 1708 RSFOKTRHFTAAVERLMDYGMSSPHVYRANAGSGVPOFKVYQOEOTDGSFPOPLY 1767
Db 1441 MAGDKNRTYFTISVEEEMDYA--GYGQRRLDTAKNNEPTYFRVAVYRRLDSTFIRDI 1499

```

```

Qy 1768 KQELNEHLGLPIYIAEVEDINWTFERNQASRPYSFYSLSIY-----EEDQROGA 1819
Db 1500 RQENDEHLGILGPLIAEVDQVWVFFNRRASRYYSILHANGKYVILKMEGLSVDSDSPW 1559
Qy 1820 EPRKNFVNPNETYTFWAKVOHMAFPKDEKQAMAYFSDVDLEKDVHSGILPLVCHT 1879
Db 1560 YKODAVAPNQTFTYMTWITNPKSGQNNESDCRTWYYSVAVNPENDINSIGLVLVCR 1619
Qy 1880 NTLN--PAHGRQYTVQEPALFTETIDEFKSYFTEPMNENCAAPONTOMEDPPEKREYF 1937
Db 1620 GLDQKPPEDR-----EYVLLMTDEKSKMLYEENRQIERKANRYVA--DPNFQDLKF 1673
Qy 1938 HAINCYIMDILPGLVAAOQRIARTLMSGSENIHSIHFSCHVTVRKKEETKALYK 1997
Db 1674 DALNG--IYLSLKGRLMYTQOLAKMHLINMGSPDLNHSVH HQDTFINKELDRHQGYPL 1732
Qy 1998 YQVEFVEMLPKRAIMRECLISEHLAAGSTLF-----LVYSNK 2039
Db 1733 LRGFATLEMLPSKPGIMQLESEVGLSOORQMOTLFLFDESYYKRYVCQVLRSLYFVAV 1792
Qy 2040 COTPLGASGHIRDFOITASGOYQOMAPKLARLHYSGSINAMST--KEPFSWIKDILAMP 2098
Db 1793 CDRPLGLISGVQDEQITASDRGQWYFHLARLHNGKYNAMSTSESPGYQVYDPOBPV 1852
Qy 2099 ITHGKIQGAKQKFPSSLYISOFILYSLDGKKNQOTYRGNSTGT-----LWVFF-----2146
Db 1853 VLSKATQGAQKQPLTNHFNVLNTTISTDARKKAYIKGSDAVRANIHKTINFFNAROT 1912
Qy 2147 --GNDSGGINHINFPPIIANYITLAPTHYSIRSLRMLGCDLNSCSPLGAMESKAI 2204
Db 1913 FEGNAEATEKENTEFEPPLIGRVYRLHLSHNPYVRLLEYGCELDGCVPLAMEGIL 1972
Qy 2205 SDAQITASSYPTNMEA--TWSPKARLHLGSHNAMPPOVNNREMLQVOPKTKRYGYT 2263
Db 1973 DSKRTASSVSNMYSQOMHPYATLINOGTANAMQAKNNDLOPIQVLELEVYKKTIGY 2032
Qy 2264 TQGVKSLTSMYKEFLISSODGQWTLFQNG--KVYFGQNDSTSTPVVNSLOPLT 2321
Db 2033 TQGAKSMGNEFVYSYILEYSEDSGRMKKRYTDDDDYQKPLFGQNTDNGQIKNTIYPPF 2092
Qy 2322 TRYLRHPOSVVHDLRMVEYGC 2346
Db 2093 SREIRLIPKQMSYVYKRIELGCD 2117

```

```

RESULT 4
ID 043737 PRELIMINARY; PRT: 2224 AA.
AC 043737;
DT 01-JUN-1998 (TREMBlrel. 06, Created)
DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE Factor V.
OS Homo sapiens (Human).
OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxId=9606;
RN 111
RP SEQUENCE FROM N.A.
RA Bird C.
RL Summited (JAN-1998) to the EMBL/Genbank/DBJ databases.
CC 1- SIMILARITY: CONTAINS 2 P3/8 TYPE C DOMAINS.
DR EMBL; 299572; CAB16748.1; -.
DR HSSP; P12259; IC2T.
DR InterPro; IPR001117; Cu-oxidase.
DR Pfam; PF00394; Cu-oxidase; 3.
DR Pfam; PF00754; F5_P8_type_C; 2.
DR SMART; SM00231; FAS8C; 2.
DR PROSITE; PS01285; FAS8C_1; 2.
DR PROSITE; PS01286; FAS8C_2; 2.
DR PROSITE; PS00079; MGLTICOPPER_OXIDASE1; 2.

```



```
Db 1009 IFR-----KKNKKIAL-----HSPSPGCDP-----1033
Qy 1125 ARMIORTGKNLSNGOGSPKOLVSLGPEKSEVQNTLEKKNVYVVGKEFTKDVGLKE 1184
Db 1034 -----LRGNHSP-----1041
Qy 1185 MYPPSSRNPLJNNDNHNHNHNOEKKIOEIEKEKTELIOENVLPOL--HYVGTGKNF 1242
Db 1042 ---FPDRRLNLSL--LHKS-----ETALSPYLNQTSMSKTDRLSDVINOQSKNDTDQM 1093
Qy 1243 MKNLFLSTKQNVGSDGAYAVYLODFRSLDSTNRKTKHHTAFKSGKEENL-----1296
Db 1094 SSSLDLY---QSVAPERSPTFPA--ODPDQHTSTDPSTYRSRSPELSOGIGLDLHDEYP 1149
Qy 1297 EGIQNOTQIWEKACTRISPTSOQNFVTOQRKALKOFRPLP--ETELEKRII---1351
Db 1150 DDIG-----LTSFPDQSKSSFSDDQALPSSDLSLFTISPELDQITIPD 1197
Qy 1352 YDDTSTQMSKMKHLPTSTLTOI---DYNEK---EKGALTOGPLSDCLTSHSIPQAN 1403
Db 1198 LDQLLSPEDNOKTSSPD--LGQVPLSPDNOKTSSPDIGQVSLSPDNOKTSS---PDLG 1253
Qy 1404 RSPLPF---AKVSPSPSIRPIYLT---RVLFQDNSSHLPASTRKKSGVQESSHTLQG 1456
Db 1254 QVPLSLDNOKTSS--PDIGQVPLSPDNQMTSPDLGQVPLSPDNOKTSSPDIGQVPLFP 1312
Qy 1457 AKKNLSIALITLLEMTGQREVGSLGTSATNSVTKKVENYLPKPD-----LPTSGKV 1511
Db 1313 EDNQNFYDLQVPLSPDNQ---ETSSDLDLTLSPDQGTVPSPDLQDLPLPSDNGOV 1368
Qy 1512 ELIPKH-----IYOKDLPTETNSNGSPHLDVYGLDGTSEALKMWEANRP 1560
Db 1369 TVSPDLSLTLSPDNELLAPDLGOVTLSP-----DLITN-----1405
Qy 1561 GKVPFLVATSESSAKTSPKLDPLANDNHYGQIPEKESQKESPEKTAFAKKDTIISL 1620
Db 1406 ---PALNHCKKASSADPDQASSTYP--DSQASSLP--ELNFTLPHRDLHIPPSPSLP 1458
Qy 1621 NACESNHAIAMINEGUNKPEIEYTWAKQRTRELCSQNPV---LKNHOREITRTLOSD 1677
Db 1459 N-----NTLSRKNPPLVYVGLSRVDDGVEIYVSEE 1490
Qy 1678 QEIDYDQTSYEMKKEDPDYEDENQSPRFOKTRH-----YFI 1719
Db 1491 PERIDED-----YAEDDFTYINDPRTDTRIDVANSNPOTIAWMLKRGHGKKFTYI 1544
Qy 1720 AAYERLMDYGMSSSPHYLRNPAOSGVPO---FKVYVQOFTDGSFTQPLIRGELNEHLG 1776
Db 1545 AAEELTWNAEFAOSEM--DHEDTGHTPKDTYKKVFRKYLDTFTSRDPAEYEEHLG 1602
Qy 1777 LIGPYIRAEVEDNIMWTFERNQASRPYSLSLISE-----EDORQAGPRKKNFYAP 1828
Db 1603 ILGPIVRAEVDVLIQRFKNLASRPYSLAHGLSYEKSEKTEYDESPWFQEDDAVOP 1662
Qy 1829 NETKTYEMKVVUOHNAPTDEDFDCKANAYSDVDLKDQVHGLIGLIPLVCHTNTLPAAGR 1888
Db 1663 NSSYTYWHAATKRSPEHPGACRAHAYYSANVERDHSGLIPRLICCKGTHMERL 1722
Qy 1889 QVTVOEFALEFTIIPETKSMFTTENNERKCARPCIOEMEDTEENTRFAINAGTYMOTL 1948
Db 1723 PMDREFFVLEWFEKKSMTYEEKS--KGR-----RLESPEENAKRFAINALGMYN-L 1774
Qy 1949 PGLVMAODORIRMYLLSGNSNENHSHSGVHTVRKKEEYKALYNLPGVFEVEML 2008
Db 1775 PGLRMYEOWEWMVHLHLNNGSGRDIHVHFGOTLDNRTKQHQGVMPRLPSGFKLEMK 1834
Qy 2009 PSKAGIIRVCECLIGHLHAGSTLFLYNSKQOTPLGASGHIDRPTGASQGYOMAPK 2068
Db 1835 ASKPGMMLDTEVGNQVAGMQTFPLIDKECKMPGSLSTVVISQIRKSEYLIWTEWR 1894
Qy 2069 LARLHAYSGSINAMSTKE-----PFS--WIKYDLLAPMIHIGIKTGAROKSSLSOFTI 2122
Db 1895 LARLNNAGSYNAMSTIEKTAIDLPFKIPWIOVMQKEVVVTGIOTOGAKHYLKSCTTEFOV 1954
```

```
Qy 2123 MYSLDGKKMOTYRGNSGTGLMVFEGNVDSGKIKNIFNPPIARTLRHPHYSTRTLR 2182
Db 1955 AYSSDOTMOTIFRGKSGKSYMFTGNSDSTIKENRDLPRIVARYIRHPKSYNRPLR 2014
Qy 2183 MELMGCDLNSCMPLGWSKSAISDAQITASSYFTNMFAT--WSPKARLHLOGRSNAMPQ 2241
Db 2015 LELQCEVNGVSGSTGLGEDRIDQKITASSPKKSMWGDYWPBSLRLMAOGRVANMAQK 2074
Qy 2242 VNNPKENLVDFQKTMVYGVYTGQYKSLTSMVYKEFLSSQDGHQWTLTFQNGKY-- 2299
Db 2075 ANNNKQMLQVPLIKKRYVATVYQGKSSKSMYKSTQYSDQGVAKKPIRKQSSWVD 2134
Qy 2300 KYFQGNODSFPPVYNSLDPPLRLRYLRHIPSOWHQAIALMEYVLCGEADILY 2351
Db 2135 KIFEGNSMTKGMMKNFNPPIISFRITPIKTMQISALRLLELFGC---DIY 2183

RESULT 6
Q92024 PRELIMINARY: PRT: 1157 AA.
AC Q92024
DT 01-MAY-1999 (TREMblrel. 10, Created)
DT 01-MAY-1999 (TREMblrel. 10, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE Hephaestin.
GN HEPH OR HEPH.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10990;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J;
RX MEDLINE=9140771; PubMed=9988272;
RA Vulpe C.D., Kuo Y.M., Murphy T.L., Cowley L., Askwith C., Libina N.,
RA Gieschler J., Anderson G.J.;
RT "Hephaestin, a ceruloplasmin homologue implicated in intestinal iron
RT transport, is defective in the sla mouse.";
RL Nat. Genet. 21:195-199(1999)
DR EMBL: AF082567; AAD16035.1; -.
DR HSP: P00450; 1KCW.
DR MGD: MGI:1332240; Heph.
DR InterPro: IPR001117; Cu-oxidase.
DR InterPro: IPR002355; MultiCu-oxidase2.
DR Pfam: PF00384; Cu-oxidase_3.
DR PROSITE: PS00079; MULTICOPPER_OXIDASE1; 3.
DR PROSITE: PS00080; MULTICOPPER_OXIDASE2; 1.
SQ SEQUENCE 1157 AA; 129662 MW; 411B1DA2BEC2FD CRC64;

Query Match 11.18; Score 1382; DB 11; Length 1157;
Best Local Similarity 20.08; Pred. No.5.3e-72;
Matches 411; Conservative 179; Mismatches 415; Indels 1052; Gaps 26;

Qy 20 ATRRYVLAGVELSDMYQSDLGELPDAFPRPVKSP---PFTNSVYKKTLTVEFTD 75
Db 24 AINRYVLAGIOMONRYAPKGRNVITQNLNDYVASSLSGKNRGRSSKTYVKEKSD 83
Qy 76 HLENIAKPPPPMGLGPTTIOAEYVTVYITLKMAHSPVSLAAGVSTKMSBGAEDD 135
Db 84 GYTERIAPPAVLGFLQDAEYGVLIILHUKNAPSPYTHIHGHGFEKSESLP 143
Qy 136 QTSOREKEDKVFPGSGSHYVVOYLKENGPNASDPLCLTYSLSHVDLYKDNSLIGAL 195
Db 144 GSSGYLAKDADSVPRGSHYVNMISIPESHAPTEADPACTIWHSHVDAPDIAITGLIPL 203
Qy 196 LVORESL---AKEKQOTLHKRTLLFAVFDGSKWSE-----TRNSLMODRBAASAR 245
Db 204 ITCKGTGLDGNSPPRKQDVNDFLPSYIDNLSMHLNDIATYCSDRASVDKEDGAFQ 263
Qy 246 AMPKMTFVGVNNSLPGGLIGKRSKSYVNHVIGMTGTPEVHSIFLEGHFTLVNHRQASL 305
Db 264 DSNRMHAINGFVGNLDELQWCAQKHYAMHLMGMCNEIDVHTAFHGMOLSTRGHHTDVA 323
```


Db 1043 MRCCHTDDHVMKMEITFTVLSHE 1065

RESULT 8

Q9B0S7 PRELIMINARY; PRT; 1158 AA.
 AC Q9B0S7;
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
 DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Hephaestin.
 GN HEPH.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Syed B.A., Beaumont N., Evans R.W., Stral S.R.;
 RT "In silico cloning and comparative modelling of human Hephaestin."
 RT Submitted (JAN-2002) to the EMBL/Genbank/DBJ databases.
 DR EMBL; AJ296162; CAC35365.2;
 SQ SEQUENCE 1158 AA; 130450 MW; A3FDBF4D35F1E5AE CRC64;

Query Match 10.8%; Score 1338; DB 4; Length 1158;
 Best Local Similarity 19.4%; Pred. No. 26-69;
 Matches 420; Conservative 199; Mismatches 449; Indels 1092; Gaps 32;

QY 20 ATRRYVLGAVELSDRYM---QSDLGELPYDAPFRPRPKSF---PNTSVYTKTLYVE 72
 DB 24 ATRRYVLGIRDVQMYAPKGRNVYITNOPLDSDI---VASSFLKSDKNRIGGTYKTKTYKE 80
 QY 73 FTDHLFNIAKPRPWWGLGPTIOAEVYDTVVITLKNASHPPVSLAAVGVSYWKASGEAE 132
 DB 81 YKDSSTDEVQAPALGFLGPIVLAEGVDVILHLKKNATRPYIHFHGVEFEKESSESL 140
 QY 133 YDQTSQREKEDKFFPGSHYVQYLKENGPMASDPLCTYSYLSHVDVLDLNSGLI 192
 DB 141 YPDGSSGPLKADSDVPGSGSHYVMTIPGCHAPTDAPACLTWYHSHVDAPRDIATGLI 200
 QY 193 GALLVCREGL---AKKQTGLHKFLLFAVFDDGKSMH-----SEFKNSLMDDRDA 242
 DB 201 GPLITCKRGLDGNSPPRODTHDPFLFVYVDENLSHNLNENIATYCCDPASVDKDE 260
 QY 243 SARAPFKMTVNGVYVNSLPELIGCHRSKYVHWITGMCTPEVHSIFLEGHTEFLVNRQ 302
 DB 261 TFQESNRMAINGFVFGNLPGLNCAQKRVANHLFGMNEIDVTAFFHGOMLTTRGHHT 320
 QY 303 ASLEISPTFLTAQTLMDLGOFLFCHISHSHQHGMEAYVYKVDSCPEEPOLMKNNBEA 362
 DB 321 DVANIFPATEFVTAEMVPEPGTMLISCOVNSHPRDGMOLAKVKVCSMAPV----- 372
 QY 363 EDYDDDLTSEMDVVRFPDDNSPSFQIISVAKKHPTKVYVIAAEEDMDVAPL----- 417
 DB 373 ----DLITG-----KRYQFIANHLIDMDYPMGHDS 401
 QY 418 ----VLAPODSYKSOYLANNGPORIGKTKYKRYMAVYDEFFKTREAIQHS--GILGPI 471
 DB 402 TGNLNEPG--SISDFQKSSSSIGGTYWKVYAFQDEFQKMLLEDRLGLIGLV 459
 QY 472 LYGEVGDLLIIFKQNASPYNIYPHGIDVPLYSRRLPKGVNHLKDFPIIGELFK-- 529
 DB 460 IRAEVGDITQVVFYRRASQPSMQPGV---FYEKDYEGTV--YNDGSSYGLVAKPF 512
 QY 530 ----YKWTVEGDTSPKCLTRRYSSPYNMEDLASLIGLILCYEESVDQNGNDI 585
 DB 513 EKYTYRMVYPPHAGTADPACLTWYFSAADPIDTNSGLVGLVLCRAGALGADGKQ 572
 QY 586 MSDKRVILFSYFDENBSWYLTENIQFLPNPAGVOL-----EDP--PQASVWHSNG 638
 DB 573 GVDKEFFLLFTVLDENKSMWNS-----ANQAAAMLDFRLSLSDIGFQDSNRMAING 625

QY 639 YTFDSL-QLSVCLHEVAYTYIISIGNQDPLSFYFSSGYTFKAKNATYEDTLFLPPSGEYV 697
 DB 626 FLFSNLPRLDKCKGDVYAMHLGLGLGETVDYGVFQGNVQLOGMKKAAMLFPHTVMA 665
 QY 698 FMSMENPGLMITLGCNSDEFNRNQMATALKVSCKDNKGDYEDSDYEDISAVILSKRNAIE 757
 DB 686 IMQPNLGLTFEITYCOAGSHREAGMAYVNSQC----- 718
 QY 758 PPSFQNSRHSSTQKQFNATTTIPENDIEKTDPMFAHRTMPRIQVNSSDILMLNQSP 817
 DB 719 -----PGHATPRQRYQARI----- 734
 QY 818 TPRGLSLDLOEATYEFSDPBPQADSNNSLSEMHFRPOLHSGDMVTPPSGQLR 877
 DB 735 ----- 734
 QY 878 LNEKLGTTAAETLKKIDFKVYSTSNNLSTIPSDNLAGTNTSLGAPSPVHYDSOLD 937
 DB 735 -----YY----- 736
 QY 938 YTLFGKSSPLTESGPLISEENNDKILSGLMSQESSMCKNVSTESGRLEFKGRA 997
 DB 737 -----IMAEVEHD----- 745
 QY 998 HGPALLTKDNLKRVSISLTKTKTSNNSATNRKTHIDPSLLIENSPEWQNLIESOTE 1057
 DB 746 ----- 745
 QY 1058 FKKYTPLIHDMLMDKNATYALRLNHSKTKTSSKNMEMVOOKKEGPIPPDAQNDPMSEFK 1117
 DB 746 -----YCPDRS-- 751
 QY 1118 MLEPESARMIQTHGKNSLNSGQSPKQVLSLGPERSYEGNPLSEKKNKVVYKGGEFT 1177
 DB 752 -----WEREHNOS-----EKDSYGIFFLSNED 774
 QY 1178 KDVGLEKAVFPSSRNLEFLTNLHENTHNOEKKIOEIEKRETLIOENVYLPQIHVT 1237
 DB 775 ----- 774
 QY 1238 GTRKPKNULFLSTRQVBSGYDGAAPVLDQFSLNDSNTRKRTKTAFAFKSGEENLE 1297
 DB 775 ----- 774
 QY 1298 GLGNQTOIVERVACTTRISPTNSQNFVQSRKALKQPLBEETELERKIIVDTST 1357
 DB 775 ----- 774
 QY 1358 QMSKNNKHLTPSTLTQIDYNEKEKGAITOSPDLQCTRSHSIPQANRSPLPYAKVSFPS 1417
 DB 775 ----- 774
 QY 1418 IRPIYITNVLFQDNSHLPAAVYRKKDSGVQESSHFLQAKKNNLSLALITLQMGQRE 1477
 DB 775 ----- 774
 QY 1478 VGSIGTSATNSVYTKKVENTVLPKPLPFTSGKVELLPKVHIYQKDLPTETSGNSPGHL 1537
 DB 775 ----- 779
 QY 1538 DLVEGSLQTEGAIKNNNEANRPKVPFLVATBSSAKTESKLDELANWHYGTQIPKE 1597
 DB 780 ----- 779
 QY 1598 EKSQEKSPKTAFAKKKDTILSLNACESHAAALINSGKPKPIETVMAKQGRTERLCSQ 1657
 DB 780 ----- 779
 QY 1658 NPVYLKRNOREITRTTLOSQBEIDYDTISVEKKKEDFDIYEDENOSPNSFOKTRHY 1717
 DB 780 ----- 779
 QY 1718 FLAAVERLMDYGSSSPHYLRNRAQSGVPOFKRVYFOEFTDGSFTQPLYRNGELNEHLGL 1777

Db	141	YPOSSSPLKADDSVPKSHIYMWLTPEBGNAPTPADPACLTUHYSHDAIDATLGI	100
OY	193	GALLVCEBGL-----AKKOTQLTKFLLFAAFEDGKSMH-----SEYKSNLOMDRPA	242
Db	201	GPILTCRGRADGNSPPRODYOVDHDFLLFVVDENLSMHNENINATYCSODSPADKDE	260
OY	243	SARWMPKMTVNGYVNRBLPGLIOCHKSYWNVHVGMTPEVHSIFLEGHIFLYNRKO	302
OY	261	TFQSNMMAHINSEFGMLPELNMCAQKRAVHLEGMONEIDVHAFEBGOMILTRGHT	320
OY	303	ASLEISPTFLAQTLLMDLGOFLTECHSISHOHDGMEAYVYKVDSCPEEPOLMKNNDEA	362
Db	321	DVANIPEPATVVAEMVPEPMTWLISCOVNSHFRGOMOLATKYKSCSMAPV-----	372
OY	363	EDYDDDLTSEMDVYRFDDDNSPSFIOIRSVAKKHPRKVVHVIJAAEBEDMDAPL---	417
Db	373	-----DLTIG-----KYQFTEIAHEDIMWDGPMGHNDS	401
OY	418	-----VLAPDRSKSOYLNNNGPORIGRKYKVFMAUYDEFFKREAIQHS--GILLPL	471
Db	402	TGNMLRRBP--SLSDKPFQKSSHGIGTGMKRYEAFODETFQKHNLEBDHNLIGLAP	459
OY	472	LXGVGQTLTIFKNOASRPVNIYHGTIDVRYLSRYLPGVCHLKDRPLRGHTEF--	529
Db	460	IRAEVGOTIOVYVYRNASOPFSMOPBY-----FEKYEESTV--YNDSSYPLGVAKRF	512
OY	530	-----YKMTVVEDGPTKSDPCLTRYSSTVNMEDLASGLIGELCTKESVDORNGOI	585
Db	513	EKVYTKMTVPPHAGPPADPACLTWMTFFSLADRIDTNSGLVGLGALCYRGAALGADQK	572
OY	586	MSDKRNVILTSVDENRSVYLTENIOFLRPNPAGVOL-----EDPE-FQASNIHMSING	638
Db	573	GVDEKFFLLTVDDEKSKYSN-----ANQAAALMFLRLSIEDGEGDSENRRHAIAG	625
OY	639	YVPSOL-OLSVCHHEVAAWYILSIGAOTDELVSFESGTYFKHKVUYEDLTFEPSCETV	697
Db	626	FLFSMLRIDLMCKGDIYAMHNLIGSTEDVHGMVQGTVOLOQMRGAAMLFPITFVMA	665
OY	698	FYSNENGLMILTICHSIDPFRNRGRTALLXYSSODKNTDYEDEYEDISAYILSKNMAL	757
Db	686	IQPDNLCTEYELCOAGSHREKGRALYVNSOC-----	718
OY	758	PRFSNSRSPSTROKQENATITPBDIEXTDPEFAHRTKPAQIONSSDILMLLRSP	817
Db	719	-----PGHOAPPRKROAARI-----	734
OY	818	TRHGSLSDLOEAKYETPESDPSGAIDSNNSLSEKTHFRPOLHNSGDMVFPESGLQLR	877
Db	735	-----	734
OY	878	LNELKGTAAATELKIDLFKVSSTSNLITIPSDNLAGDTNITSGLPSPMVHYDSOLD	937
Db	735	-----YU-----	736
OY	938	TTLFEKSSPLTSGGRLSLSENNDSKLLSEGLMNSOESSGKNVSGSTESGRLEFKRA	997
Db	737	-----JAAEBEEMD-----	745
OY	998	HOPALLTKDNALFKVSIILKTNTSNNSATNKHITHDGPILLIENSPTWONLIESOTE	1057
Db	746	-----	745
OY	1058	FKKVTPLIHRMLMDKNATATLNMHNSKTTSSKNMENVQKKGGIIPDAONPMKSEFK	1117
Db	746	-----YCPORS-----	751
OY	1118	MUFLPESARMIORTHGKNSLNGOOPSPKQVSLAGEPKSVGEONFLSEKKNVVGGEFT	1177
Db	752	-----WEREMHNS-----EKDSYGIYFLSNRQ-----	774
OY	1178	KOVGLAKEMVPPSSRNFLTINDNLHENNTNOEKKIQEIEIEKKEFLIQEYVVLQIHTYT	1237

[illegible]

Query Match	10.4%	Score 1292;	DB 11;	Length 1084;
Best Local Similarity	19.2%	Pred. No. 8.7e-67;		
Matches 397;	Conservative 198;	Mismatches 423;	Indels 1054;	Gaps 29;
OC	Mammalia; Euthertia; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.			
OK	NCBI_TaxId=10116;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=SPRAGUE DAWLEY; TISSUE=BRAIN;			
RX	MEDLINE=20127919; PubMed=10660599;			
RA	Patel B.N., Dunn R.J., David S.;			
RT	"Alternative RNA Splicing Generates a Glycosylphosphatidylinositol-			
RL	anchored Form of Ceruloplasmin in Mammalian Brain.";			
RJ	J. Biol. Chem. 275:4305-4310(2000).			
DR	EMBL: AF202115; AAF34175.1; -			
DR	HSSP: P00450; ICKG.			
DR	InterPro: IPR001117; Cu-oxidase.			
DR	InterPro: IPR002355; Multiclu_oxidase2.			
DR	Pfam: PF00394; Cu-oxidase; 3.			
DR	PROSITE: PS00079; MULTICOPPER_OXIDASE1; 3.			
DR	PROSITE: PS00080; MULTICOPPER_OXIDASE2; 1.			
SO	SEQUENCE 1084 AA; 123749 MW; 9506584243257055 CRC64;			
QY	5 LSTGCFLLRPFCSATRRYYGAVELSMQMO-SPLGLL-PVDAARPPRRVSPSPFNTS	62		
DB	6 LSAALLF-HSSLAMTREKHYYIGILEAWMOVAGSSSEKRLIVDEQSNLYLNGDRIG	64		
QY	63 VYKRTTLVETDHLFNIAKPPRPMLGLTQIAVYDTVYITLKNAHSPPSLAVAY	122		
DB	65 RYKKAALSETTDITTKTIDPAAVLGAVLKAEGKSVANAKNAPSPTFAHAY	124		
QY	123 SYWKASEAEYDQTSOREKEDKYFPGSHYYWOYLKENGPMASDPCLITYSHVD	182		
DB	125 YTTKAMBSAIPYDNTWTFDRADKDLFPEQOYLVLVA-NEPSGSGDSNCTRYTSHVD	183		
QY	183 LVKDLNSLIGALLVREGSLAKERTQTL-HKFIILFAVFDGSKMSHETKNSLM----	236		
DB	184 AKRDIASGLIOPRLTKKSGSLKREKEENIDDEFLVMEVYDENISWYLEDNIKTFCSPE	243		
QY	237 -ODRPAASARAPMHTVNGVYNRSLPGLGCHRSKYHNVHGNQTPPEVNSIFLGHPF	295		
DB	244 KYDKDNEDFOFSNNRMTSINGYFGSLPGLSKCAEDRYKYLFGMGENVDVSELFGQAL	303		
QY	296 LVNRQASLEISPTFFLAOTLMDLQDLFCFCHSSHOHGMGAAYKVDSCPEEPPQR	355		
DB	304 TSKNYHTDILNLPATLIDVSWVAQNPQVWMLSGQNMLHKKAGLQAFQVWDC-----	356		
QY	356 KMNNEADYDDDLTDESDVYRFPDDNSPFIQISVAKKRPKTVHNYIAAEEDMDVA	415		
DB	357 ---NKPSP--DDDLQDNH-----RH-----YIAAEETLMDA	385		
QY	416 P-----LVLAEDRSKYSOYLNNGFORIGRKYKRVPMAYDTEF--KTREAI	461		
DB	386 PGDTFTFGENTSLSGDSVFEEQ---GATRIIGSKYLLVREYTDOSTNNKRGPD	441		
QY	462 OHBSGILPRLYGVGVTLLIIFKQAQSRPNVYIPIHSINDV---PYLSRLPKQKHL	517		
DB	442 EEHLGILPVIAMVEGDIIRVTFHNKQGFPLSIOPMGVRFREKNGTYG---PDGRSSK	498		
QY	518 KQPLILPEIKRYKWTYVDEGPTKSDRCGLTRYSFNNMRDLASGLIOPLLICYSKS	577		
DB	499 QASHVAPEFTTYMYIYKREMGPTIADVCYSKRYISGVDLKDLPFLLIOPMKLCKGS	558		
QY	578 VDRGNQMSKRNVLISYFDENBSWLTENIORFLPNPAGVQULEDEPQASINAHIN	637		
DB	559 LLAADGQADVKEFLYAVYFDENESLLLDNNIMFPLVAENVEDDEDPESSKMSNKA	618		
QY	638 GAVFDSLO-LSCVLEHVAWYIISIAQDTFLSVFSGYTFKHKWYEDTLTLPFESGET	696		
DB	619 GPWYGLNGLGIMCKGSEIWAYLFSAGNADVHGVIYFGSNTIYLSKEBDRTNLNLPKHSIL	678		
QY	697 YFMSKDNGLIILGCHNSDPRNKGATALKKYSQCDKNTQGYDEYDEYISVLLSKNNAI	756		

Db 300 HGOVLSKRYVDITNLFPAITLFDIAEVAONQOMLSCQNLHLKAGLAFWVODCK 359
Qy 351 EPOLRRKNNNEAEDDDLDSEMDVVRPDDDNPSFIOIRSVAKKHPTVHIAAEE 410
Db 360 SS-----YIAAEV 381
Qy 411 DMDYAPL-----VLAPDSSYKSOYLANNPORIGKRYKRYKFAAYIDEP---KTR 458
Db 382 INMYAPSGIDAPYKENLAPESAS--EAPFEGQPTIGSYKKLVARYEDASQOKER 439
Qy 459 EAIQESGILGPLYGEGDTLLIFKNOASRPYNYIPGI-----TDVAPLSRRL 510
Db 440 GPEEHLGILGPIVIAAEGDITIRTFENKRAHPLSTIEPIGVYDKNNGEYTPSGSRP 499
Qy 511 PKGVKHLKDFPLPGELIFKRYKRYKRYVEDGPTSDRCLETRYSSVFMEDLASGLIGL 570
Db 500 PSG-SH-----VAPRGFTYEMTPKREVGPTYKDPVCLAKMYG--GSTFDITGLIGPM 551
Qy 571 LICYESVDQKQNOIMSDKRNVLFSVFDENSWYLTENIQRFPLNPAGVQLEDPEFOAS 630
Db 552 KICRNGSILANGRLKNDKFEYLFPTVFDENSLLIDNIMKFTTAPDQYDKENEDPQES 611
Qy 631 NIMHSINGVFPDSLO-LSVCLHEVAWYLIISGAGTDFLSVFPESGYTFKKMYEDTTL 689
Db 612 NKMHSNMGFTMGNOPLSKOGDSVWMTLFSAGNEVDHIGYFSGNTYLSRGERRDTANL 671
Qy 690 FPGSGEYVPMGMEPMGLMILGCHNSDFRNKMTALLKYSQDKMTGDIYEDSYEDISAVL 749
Db 672 FPQTSLSLEMQPDAGFDECLTTHYGGMKQRYTVSQCQGSSELY----- 720
Qy 750 LSKRNAIEPRSFQNSRHPSTROKOPNATTIPENDIEKTDPMFAHRTPMKIONVSSDL 809
Db 721 ----- 720
Qy 810 LMLNQSPTRPGLSLSDQAKYETSDPSGALDSNNSISEWTHRPOLHHSQMTV 869
Db 721 ----- 720
Qy 870 PEGLOLRNLEKLTATATLKLDFKVSSTSNLITIPSDNLACTDNTSISLGPSPW 929
Db 721 -----LYLGRTRYIAAVE----- 736
Qy 930 VHYDSQDITLFGKXSPPLFESGSPLSISENNDSKLLFSGIANQSBSGKXVSTESG 989
Db 737 ----- 736
Qy 990 RLKGRAGHPALLTKDNALFKVYISLTKNTSNNSATNRKTHIIDPSLLIENSPVWQ 1049
Db 737 -----WD 738
Qy 1050 NILESDFEKKVPTLIHDKMLDKNATALRLNHSNKTTSKKNEMVQOKKSGPIPPDQ 1109
Db 739 ----- 738
Qy 1110 NEDNSFFKMLFPESARWIORTHGKNSLNSGQSPKQVYSLAPEKSEVQONLSEKNV 1169
Db 739 ----- 738
Qy 1170 VVGKEFTKIDVGLKEMVFPSSRNILFTLNDLNHNHNHNOEKIOEIEKKEILLQENV 1229
Db 739 ----- 738
Qy 1230 LPQIHVYGTGRNPKMLFLISTRONVSGSYDGAAPVLODFRSLINDNTRKTHAFHSK 1289
Db 739 -----YSP----- 741
Qy 1290 KGEENLEGLOTKOIVEACTTRISPNTSQNFYQSKRALKQFLPLEETLEKR 1349
Db 742 ----- 741
Qy 1350 IIVDDTQNSKMKMLTSPSTLQIDYNEKEKGAITQSPSLDCLTRSHSIPQANRSPPI 1409
Db 742 -----SRKWEKELHH----- 752

Qy 1410 AKVSFSPISRPILYLRVLPQDNSSHLPAASYRKXKQSGVSSHFLQAKNNSLAILTL 1469
Db 753 -----OBONISNAFL-- 762
Qy 1470 EMTGQREVSLGTSATNSVYKRYKENTVLPKPDLPKTSKVELLPVHLYORDLEPET 1529
Db 763 ----- 762
Qy 1530 SNGSPGHLIVESLQGTGCAIKKMEANRPQVPLRVATESAKTPSKLLDPLANDH 1589
Db 763 ----- 762
Qy 1590 YGTQIPKEWKSQEKSEPTAEKKDTLLSLNAGESNHAIANEGONPELEVWAKG 1649
Db 763 ----- 762
Qy 1650 RTERLCSQNPVYLKRQRETRITTLQSQOEIDYDITVEMKKEPDFDIYDEDNQSPRS 1709
Db 763 -----DKEEPTI-- 769
Qy 1710 FOKTRHYFLAAVERLMDYGMSSPHVLRNRAQSGSVPOKRYVFOETFDGSPFQPLRG 1769
Db 770 -----GS--KRYKRYKQFTDSTFOYVPER 793
Qy 1770 ELNEHLGLGPIYRAEYEDINIMVTFRNOASRPYSLSLSTYEDDQROGAERPKRYPN 1829
Db 794 GEEHILGILGPOLHADVGDKNVILFRKNATRPYSIHAQVKTSSST---VTP---TAPG 846
Qy 1830 EKTTFYKVOHHAFTKDEFOCKANAYISVDLEKDVHSLGILPLVCHTNTL---NPAH 1886
Db 847 ERTYIMKIPERSGAGNGSPGIPWYVSTVDRYKDLFSLIGLIPLYCKRHLYKSNP-- 904
Qy 1887 GROVYOEALFETIDETKRYWTEENNERNCRAPCINIOEDPTFKENTRPAINGYIND 1946
Db 905 ---IKKLEFSLFLVFDENESWYLDNITKTSYSDPEKVDANEFMESKHAINGREF 961
Qy 1947 TLPGVMAQQRIRRWLLGMSNENIHSHGNYFVRYKKEEYKALNLYPVFEVTE 2006
Db 962 NLOGLTM-----HVGNEVDLSVHFRHGSFQYORHGYTSVDPLDPGYOTLE 1010
Qy 2007 MLPKAGIMRECLIGEHLLHAKNSTFLVYSNK 2039
Db 1011 MTPKTPGIMLLHCHVTVDHIAHMETTYVLVNE 1043

RESULT 12
Q14286
ID Q14286 PRELIMINARY; PRT; 216 AA.
AC Q14286;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 01, Last annotation update)
DE Coagulation factor VIII associated protein B (Coagulation factor VIII,
DE prococagulant component) (Hemophilia A).
GN F8B.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID:9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93052386; PubMed=1427887;
RA Levinson B., Kenrick S., Gamel P., Fisher K., Gitschler J.,
RT Evidence for a third transcript from the human factor VIII gene.*;
RN Genomics 14:585-589(1992).
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN;
RA Strausberg R.,
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; M90707; AA58466.1; -
DR EMBL; BC022513; AAH22513.1; -.

DR HSP; P00451; ICFG.
 DR InterPro: IPR000421; FAS8_C.
 DR Pfam: PF00754; F5_F8_type_C_2.
 DR SMART: SM00231; FAS8C_1.
 DR PROSITE: PS01285; FAS8C_1; UNKNOWN_1.
 DR PROSITE: PS01286; FAS8C_2; 2.
 SQ SEQUENCE 216 AA; 24641 MW; 6C8204F89E35A376 CRC64;

Query Match 8.9%; Score 1104; DB 4; Length 216;
 Best Local Similarity 100.0%; Pred. No. 6, 9e-57;
 Matches 208; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2144 VFGNDVSSGIRKININPIIARIYRIHPHYSTRIRLMEIAGDINSCSMPLEGESKA 2203
 DB 9 VFGNDVSSGIRKININPIIARIYRIHPHYSTRIRLMEIAGDINSCSMPLEGESKA 68
 QY 2204 ISDAQITASSYFTNMFATSPSKARLHOGKSNAMPQVNNPREMLQVDFQTKKTYGYT 2263
 DB 69 ISDAQITASSYFTNMFATSPSKARLHOGKSNAMPQVNNPREMLQVDFQTKKTYGYT 128
 QY 2264 TGVKSLTSMYKEFLISSQDGHQWTLFQNGKVKVFGQNDSTFPVNSIDPPLLTR 2323
 DB 129 TGVKSLTSMYKEFLISSQDGHQWTLFQNGKVKVFGQNDSTFPVNSIDPPLLTR 188
 QY 2324 YLRHPQSWVHQIALMEVLCGEADLY 2351
 DB 189 YLRHPQSWVHQIALMEVLCGEADLY 216

RESULT 13
 075659 PRELIMINARY; PRT; 782 AA.
 AC 075659;
 DT 01-NOV-1998 (TrEMBLrel. 08, Created)
 DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE DJ46618.1 (Coagulation factor V (activated protein C cofactor),
 DE coagulation factor VIII (Procoagulant component) and ceruloplasmin
 DE (EC 1.16.3.1, ferroxidase) like) (Fragment).
 GN DJ46618.1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Wilkinson J.;
 RL Submitted (JUL-1998) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AL030998; CAAL19742.1; -.
 DR HSP; P00450; IKCW
 DR InterPro: IPR001117; Cu-oxidase.
 DR Pfam: PF00394; Cu-oxidase; 2.
 DR PROSITE: PS00079; MULTICOPPER_OXIDASE1; UNKNOWN_2.
 KW Coagulation.
 FT NON_TER 1
 FT NON_TER 782
 SQ SEQUENCE 782 AA; 87939 MW; 2FBD2824CC19D2AB CRC64;

Query Match 8.8%; Score 1091; DB 4; Length 782;
 Best Local Similarity 35.4%; Pred. No. 2, 9e-55;
 Matches 247; Conservative 104; Mismatches 252; Indels 94; Gaps 14;

QY 65 YKTLFVETDHLFNLAKPRPMMGLIPTQAEVYTVITLKNASHPVSLHAGVSY 124
 DB 13 YKTLFVETDHLFNLAKPRPMMGLIPTQAEVYTVITLKNASHPVSLHAGVSY 72
 QY 125 WKASGAGYDQTSQRKEDKVFPGSGSHYVWQVLEKNGPMASDPLCTFYSLSHVDLY 184
 DB 73 EKDSGSLYPPGSSGPKADDSVPGSGSHYVWQVLEKNGPMASDPLCTFYSLSHVDLY 132
 QY 185 KDINSGILGALVREGSLAKERTQTLNFKLILFAVDEGKSMH-----SETKNSLMQD 238
 DB 133 RDIATGLIGLITCKRG-----TDHDFILFSVDENLSMHLNENIATYCSDPASVD 184

QY 239 RDAASARAMPKMTVNGVYVNRSLPGLIGCHRSKYVWAVIGCTTPPEVHSTFLBGFPLVR 298
 DB 185 KEDFEQESNRMHAINFVGNLPELNMCAQKRAVAMHIFGSGNIDVHAFRGMQLTTR 244
 QY 299 NHRQASLEIPITFLAQTLLMDLQGFILFCHSISSHQDGEAVYVWQVLEKNGPMASDPLCTFYSLSHVDLY 358
 DB 245 GHHTDVANIPFATVEMVPMERGWILSCQVNHFRDQNALYKVKSCSNAPV----- 300
 QY 359 NEBAEDYDDITLSEMDVYFDDNSPSTQIRSAKHPKTYVHYIAAEEDNDYAPL- 417
 DB 301 -----DLITG-----KVQYFLEAHEIQWDYGPWG 325
 QY 418 -----VLAPDRSYKSOYLNNCPQIRGKRYKRPVMAVYDTEFKTRALQNES--GI 467
 DB 326 HGGSTGKMLREBQ--SLSDKFPQSSRSYGYVWVREAVQDETFOEKHLEEDRHIL 383
 QY 468 LGPILXGEGDTLLIRKQASRPVYVHGTIDVRLPLSKRLPKYKHLKDPILPGEI 527
 DB 384 LGPILXGEGDTLLIRKQASRPVYVHGTIDVRLPLSKRLPKYKHLKDPILPGEI 436
 QY 528 FK-----YKMTVVEDGPTKSPDLITRYSSFYVMEBDIASGLIPLLICYKESVDOR 581
 DB 437 AKPEKVTYKMTVPFHAGPTAOPACTMMYTSADPIROTNSGLVGPLLYCAAGALGAD 496
 QY 582 GNCINSDKRNVLFSVDENRSTLTENQRLPMPAGYOL-----EDPE-FOASNIWH 634
 DB 497 GKQGVKDFEFLFTLVLDENKSWYSN-----ANQAAAMLDFLLSEDIEGFQDSNRH 549
 QY 635 SINGVYFDSL-OLASVCHHEVAYWYIISAGQTDLFVSFSGYTKHMYEDTLTLPFS 693
 DB 550 AINGFLFSNLPRLMCKGDIYVAMHLLGLGTFENDVAGVMEQNTVQLOGMKGAAMLPRT 609
 QY 694 GETVPMSENPCMLTLCGHSNDRNKGATLAKYSSC 730
 DB 610 FYVAINQPDNLGTFEITCOAGSHREGAMRAIYVNSQC 646

RESULT 14
 075180 PRELIMINARY; PRT; 891 AA.
 AC 075180;
 DT 01-NOV-1998 (TrEMBLrel. 08, Created)
 DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE KIA0698 protein.
 GN KIA0698.
 OS Homo sapiens (human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA TISSUE-BRAIN;
 RC TISSUE-BRAIN;
 RX MEDLINE=98403880; PubMed=9734811;
 RA Ishikawa K., Nagase T., Suyama M., Miyajima N., Tanaka A., Kotani H.,
 RA Nomura N., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. X.
 RT The complete sequences of 100 new cDNA clones from brain which can
 RT code for large proteins in vitro.";
 RL DNA Res. 5:169-176(1998).
 DR EMBL: AB014598; BAA31673.1; -.
 DR HSP; P00450; IKCW.
 DR InterPro: IPR001117; Cu-oxidase.
 DR InterPro: IPR002355; Multicopper oxidase2.
 DR Pfam: PF00394; Cu-oxidase; 3.
 DR PROSITE: PS00079; MULTICOPPER_OXIDASE1; 3.
 DR PROSITE: PS00080; MULTICOPPER_OXIDASE2; 1.
 SQ SEQUENCE 891 AA; 100624 MW; AB70D0436109C5B2 CRC64;

Query Match 8.6%; Score 1068.5; DB 4; Length 891;
 Best Local Similarity 31.9%; Pred. No. 7, 4e-54;
 Matches 253; Conservative 124; Mismatches 288; Indels 127; Gaps 19;

